

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2002, 15:01:41 ; Search time 26.64 Seconds  
(without alignments)  
1915.784 Million cell updates/sec

Title: US-09-714-882-2  
Perfect score: 3570  
Sequence: 1 MPELSLLIEILILGVTKT.....LFGVLIPLVLLLRNHHG 689

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3570	100.0	689	22	AAE02430 Novel human protei
2	3554.5	99.6	688	22	AAE02431 Novel human protei
3	2813	78.8	590	22	AAE02432 Novel human protei
4	2792.5	78.2	576	22	AAE02435 Novel human protei
5	2777	77.8	575	22	AAE02436 Novel human protei
6	2392	67.0	499	22	AAE02434 Novel human protei
7	2130	59.7	418	22	AAE02433 Novel human protei
8	1650	46.2	794	20	AA17750 Human pancreas-spe
9	1650	46.2	794	21	AAB41251 Human ORFX ORF1015
10	1646	46.1	766	20	AA18096 Full length mouse
11	1646	46.1	788	20	AA18095 Full length mouse

12	1150.5	32.2	574	20	AA18094 Partial Sel-1L pro
13	1098	30.8	404	20	AA18097 Human Sel-1L prote
14	435	12.2	640	21	AA195006 Human secreted pro
15	435	12.2	979	22	AA139116 Human polypeptide
16	435	12.2	1038	22	AA140902 Human polypeptide
17	373	10.4	833	18	AA127148 HMG-CoA reductase
18	300	8.4	301	21	AA166654 Membrane-bound pro
19	300	8.4	301	22	AA187534 Human PRO1063. Ho
20	300	8.4	301	22	AA165177 Human PRO4063 (UNQ
21	275	7.7	384	21	AA129891 Human secreted pro
22	275	7.7	384	21	AA129892 Human secreted pro
23	272	7.6	274	18	AA120582 H. pylori secreted
24	271.5	7.6	256	17	AA105197 Helicobacter pylori
25	268	7.5	54	22	AA120409 Peptide #6843 enco
26	268	7.5	54	22	AA134894 Peptide #8931 enco
27	267.5	7.5	256	19	AA110987 H. pylori ORF 06ee
28	267.5	7.5	256	19	AA109870 Antigen from clust
29	258.5	7.2	256	19	AA198350 H. pylori GHP0 646
30	258.5	7.2	256	19	AA198350 Expressed antigen
31	248	6.9	338	20	AA189928 Antigen 1 from clu
32	248	6.9	353	20	AA189824 Protein encoded by
33	248	6.9	355	19	AA110975 H. pylori ORF hp5p
34	238.5	6.7	185	18	AA120262 H. pylori secreted
35	238.5	6.7	185	18	AA124622 H. pylori secreted
36	226.5	6.3	290	18	AA155539 H. pylori ORF 02ge
37	219	6.1	285	21	AA120876 Arabidopsis thalia
38	219	6.1	285	21	AA153482 Arabidopsis thalia
39	219	6.1	351	21	AA120875 Arabidopsis thalia
40	219	6.1	351	21	AA153481 Arabidopsis thalia
41	219	6.1	356	21	AA120874 Arabidopsis thalia
42	219	6.1	356	21	AA153480 Arabidopsis thalia
43	218.5	6.1	291	20	AA189880 Antigen 1 from clu
44	218.5	6.1	291	20	AA189827 Protein encoded by
45	217.5	6.1	313	18	AA155443 H. pylori protein

#### ALIGNMENTS

RESULT 1  
AAE02430  
ID AAE02430 standard; Protein; 689 AA.  
XX  
AC AAE02430;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
XX Novel human protein (NHP) #1, sharing similarity with Notch ligand.  
DE Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
KW cholesterol metabolism; coronary artery disease; gene therapy;  
KW cerebroprotective.  
XX Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 393 /note= "Encoded by RAA"

FT WO200136636-A2.  
XX  
PD 25-MAY-2001.  
XX  
PF 16-NOV-2000; 2000WO-US31373.  
XX  
PR 17-NOV-1999; 99US-0165959.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX

PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
XX WPI; 2001-355635/37.  
DR N-PSDB; AAD06374.  
XX Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX  
XX Claim 2; Page 27-28; 39pp; English.  
XX The present sequence is a novel human protein (NHP) which share  
CC structural similarity with animal Notch ligands, particularly SEL-1.  
CC SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 689 AA;

Query Match 100.08; Score 3570; DB 22; Length 689;  
Best Local Similarity 100.08; Pred. No. 2.1e-298;  
Matches 689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKPLSLLEILLIILGVTIKTKAEHNKRQKERNVTVQVNEIKQYLSHILEQRTSSNV 60  
DB 1 mkplslleillilgvtiktkaehnrkrqernvttqvsneikqylshileqrtssnv 60  
QY 61 INKRENLEKKKNQKRIKIGTQNKDILKRNKNHLOKAEKNFTDEGDOLFKMGIKVLOQ 120  
DB 61 Inkrenllekkknqrirkigtknkdilkrnknhlokaeknftdegdqlfmgikvllqg 120  
QY 121 SKSQKQKEAYLLFAKADMGNLKAMEKADALLFCNFCVQNTAAIOLYESLAKEGSCK 180  
DB 121 ssksqkqkeayllfakadmgnlkamekadamallfngfcvqntaailgylslakegsc 180  
QY 181 AQNALGFLSSYIGMEYDQAKALIYYTFSGAGNMMSQMLGYRYLSGTVNLQNCVALS 240  
DB 181 aqnalgflssyigmeydqakaliyytfsgagnmmsqmlgyrylsgtvnlqncvals 240  
QY 241 YKKKVADYIADTFEKSQGVPEKVKRLTERPENLSNSEILDWDIYQYKFLAERGDDVQIQ 300  
DB 241 ykkkvadyiadtfeksegvpvkekvrleterpensnselldwdiyykflaergdvgiq 300  
QY 301 VSLGOLHLIGRGLDDYKALHYFLKAAKAGSANAMAFICKMYLEGNAVPPQNNATAPK 360  
DB 301 vslgolhligrklddykalyflkkaagksanamafickmylegnavppqnnatafk 360  
QY 361 YFSMAASKGNATLHGLGLLYPHGKGVPLNTYAEALKYFOKAEKQWPAQFQOLGPMYYSG 420  
DB 361 yfsmaaskgnatlhglglyphgkgvplntyaealkyfokaekqwpdaqfqlgfmmyysg 420  
QY 421 SGIKWDYKLAIFYLASQSGOPLAIYYLAKMYATGTGTVRSCRTAVELYKGVCBLGHW 480  
DB 421 sgikwdyklafyiflasqsgoplaiyyalakmyatgtgvtvrrscrtavelykgvcblghw 480  
QY 481 EKFLTAYAYKDGDDTSSILVQYALLAEMGYEVAQNSAFILESKANILEKMYPMALL 540  
DB 481 ekfltayaykdgddtssilvqyallaemgyevaqsnsafileskanilekmypmall 540

QY 541 LWNRAAIOGNAPARVIGDYHYGYGTTKDYOTAATHYSIAANKYHNAQAMENLAYMYEH 600  
DB 541 lwnraaiognafarvkgdyhygygtygtkdyotaathysiaankaahnaqamfnlaymyeh 600  
QY 601 GLGITKDIHLARRLYDMAAQTSDDAHIPVLFVAVMKLETHLLRLDILFFNQFTRNNWKL 660  
DB 601 glgiktdihlarlydmaaqtsddahipvlfavmkletthllrdilffnqftrnnwkl 660  
QY 661 DNTIGPHWDLFVIGLIVPGILLILNNHHG 689  
DB 661 dntigphwldfvlglivpgillilnnhhg 689

RESULT 2  
RAE02431  
ID AAE02431 standard; Protein: 688 AA.  
AC AAE02431;  
XX 10-AUG-2001 (first entry)  
DT Novel human protein (NHP) #2, sharing similarity with Notch ligand.  
XX Human; nootropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
KW cholesterol metabolism; coronary artery disease; gene therapy;  
KW cerebroprotective.  
XX Homo sapiens.  
OS  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 393  
FT /note= "Encoded by RAA"  
XX WO200136636-A2.  
XX 25-MAY-2001.  
XX 16-NOV-2000; 2000WO-US31373.  
XX 17-NOV-1999; 99US-0165959.  
XX (LEXI-) LEXICON GENETICS INC.  
XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
XX WPI; 2001-355635/37.  
XX N-PSDB; AAD06375.  
XX Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX  
XX Claim 3; Page 29-30; 39pp; English.  
XX The present sequence is a novel human protein (NHP) which share  
CC structural similarity with animal Notch ligands, particularly SEL-1.  
CC SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or

CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SO Sequence 688 AA;

Query Match 99.68; Score 3554.5; DB 22; Length 688;  
Best Local Similarity 99.98; Pred. No. 4.6e-297;  
Matches 688; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MKPLSLLEILLIGVTIKTIAEHNKRQERNVTTQVSVNEIKOYLSHILEQRTSSNV 60  
Db 1 mkplsllleilligvtiktiaeahnkrqernvttqsvneikqylshileqrtssnv 60

Qy 61 INKRENLEKKNQKIRIKIGIONKDIILKRNKHLOKAEKNFTDEGDLFMGKIVLQO 120  
Db 61 inkrenllekknqkirkigiqnkdilkrnkhilqkaeknftdegdlfmgikvllq 120

Qy 121 SKSOKKEEAYLLFPAKADMGNLKAMEKMDALLFGNFGVONITAAIOLYESLAKEGSK 180  
Db 121 sksqkqeeayllfakaadmgnlkamekmdallfgnfgvgnitaaolyleslakegsk 180

Qy 181 AQNALGFLSSYXGIGMEYDOAKALIYYTFGSAGGNMMSQMLGYRISINVLQNCVLA 240  
Db 181 aqnalgflssyigmeydakaliyytfgsaggnmmsqmlgyrlysginvlqncevals 240

Qy 241 YKKVADYIADTFEKGSEGVPEKVKLTERPENLSNSELIDWDIYQYKFLAERGDVQIQ 300  
Db 241 ykkvadyiadtfeksegvpvpekvrlterpenlsnselildwdiyqykflaergdvqiq 300

Qy 301 VSLGOLHLIGRGLDQDYKALHYFLKAAKAGSANAMAFICKMYLEGNAAVPONNATAFK 360  
Db 301 vslgqlhligrkgldqdykalyhflkaakagsanamafickmylegnaavpnnatafk 360

Qy 361 YFSAASKGNAIGLHGLGLYFHGKGVPLNYAEALKYFQKAAEKGPDAQFQGLFMYYSG 420  
Db 361 yfsmaaskgnaighlglyfhgkgvplnyaealkyfqkaaeagwpdaqfqlfmyysg 420

Qy 421 SGIKDYKLFKFFYLASQSQPLAIYLYLAKMYATGTGTVWVSCRTAVELYKVCBELGHW 480  
Db 421 sgikdyklfkyfyllasqsqplaiyllylakmyatgtgvtvscrtavelykgvbelghw 480

Qy 481 EKFLTAYFAYKDGDIIDSLVOYALLAEMGYEVAQNSAFILESKANILEKEMYPMALL 540  
Db 481 ekfltayfaykdgdiiidslvoyallaemgyevaqsnsafileskkanilekemypmall 540

Qy 541 LWNRAAIOGNAPARVIGDYHYGYTKKDYOTAATHYSIAANKYHNAQAMFNLAYMYEH 600  
Db 541 lwnraaiognafarvkgidyhygytkkdykdyotaathysiaankyhnaqamfnlaymyeh 600

Qy 601 GIGITKDIHLARLYDMAQSPDAHIPVLFVAVMKLETHLLRDILFNFQFTRWNWKL 660  
Db 601 gigitkdihlarryldmaaqspdahipvlfvavmklethllrdilfn-ftrwnwkl 659

Qy 661 DNTIGPHWDLFVIGLIVPGLILLRNHNG 689  
Db 660 dntigphwdlflviglivpglillrnnhg 688

RESULT 3  
AAE02432  
ID AAE02432 standard; Protein; 590 AA.  
XX  
AC AAE02432;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Novel human protein (NHP) #3, sharing similarity with Notch ligand.  
XX Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;

novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
Parkinson's disease; stroke; vascular dementia; fat metabolism;  
cholesterol metabolism; coronary artery disease; gene therapy;  
cerebroprotective.  
XX  
OS Homo sapiens.

Key Location/Qualifiers  
FT MISC-difference 393 /note= "Encoded by RAA"  
FT  
XX  
PN WO200136636-A2.  
PD 25-MAY-2001.  
XX  
PF 16-NOV-2000; 2000WO-US31373.  
XX  
PR 17-NOV-1999; 99US-0165959.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
XX  
XX WPI; 2001-355635/37.  
DR N-PSDB; AAD06376.  
XX  
PT Novel isolated human polynucleotides encoding polypeptides sharing  
sequence similarity with mammalian SEL-1 proteins which are negative  
regulators of Notch family receptors, useful for treating diabetes,  
cancer  
XX  
XX Claim 4; Page 31-32; 39pp; English.

PS The present sequence is a novel human protein (NHP) which share  
structural similarity with animal Notch ligands, particularly SEL-1.  
CC SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy or  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SO Sequence 590 AA;

Query Match 78.8%; Score 2813; DB 22; Length 590;  
Best Local Similarity 99.1%; Pred. No. 2.3e-233;  
Matches 549; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKPLSLLEILLIGVTIKTIAEHNKRQERNVTTQVSVNEIKOYLSHILEQRTSSNV 60  
Db 1 mkplsllleilligvtiktiaeahnkrqernvttqsvneikqylshileqrtssnv 60

Qy 61 INKRENLEKKNQKIRIKIGIONKDIILKRNKHLOKAEKNFTDEGDLFMGKIVLQO 120  
Db 61 inkrenllekknqkirkigiqnkdilkrnkhilqkaeknftdegdlfmgikvllq 120

Qy 121 SKSOKKEEAYLLFPAKADMGNLKAMEKMDALLFGNFGVONITAAIOLYESLAKEGSK 180  
Db 121 sksqkqeeayllfakaadmgnlkamekmdallfgnfgvgnitaaolyleslakegsk 180

QY 181 AONALGFLSSYIGMEYDOAKALIIYTTGSGAGNMMSOMILGYRYLSGINVLQNCCEVALS 240  
Db 181 aqnalglfssyigmeYdqakaliYttfgsagnmmsqmilgyrylsginvlqncevals 240  
QY 241 YKKKVDYIADTFEKGSEGVPEKVRTERPENLSSNSETLDDWIIYQYKFLAERGDVQIQ 300  
Db 241 ykkkvadyiadtfeKsegvpekvrtterpenlssnselldwiiyqykflaergdvqilq 300  
QY 301 VSLGOLHLITGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVPQNNATAFK 360  
Db 301 vslgqlhligrkgldqdykalyhflkaakagsanamafigmylegnaavpqnntafk 360  
QY 361 YFSMAASKGNAIGLHGLLPHGKGVPLNYAEALKYQKAEEKGWPDAQFOLGPMYYSYG 420  
Db 361 yfsmaaskgnaiglhlgllyfhgkgvplnyaealkyfqkaeekgwpdaqfqlgfmmyysg 420  
QY 421 SGIKWDYKLAIFYLASOGOPALYIYLAQMYATGTGVVRSCTAVELYKGCVELGHWA 480  
Db 421 sgikwdyklafyylasogoplaIyylakmyatgtgvvrsctavelykgvcelghwa 480  
QY 481 EKFLTAYFAYKDGDDISSLVQYALLAEMGYEVAQNSAFILSKKANILEKEKMPMALL 540  
Db 481 ekfltayfaykdgdiddsslvqyalllaemgyevagqnsafilskkanilekempmall 540  
QY 541 LWNRAAIOGNAPAR 554  
Db 541 lwnraaighslgq 554  
RESULT 4  
ID AAE02435 standard; Protein; 576 AA.  
AC AAE02435;  
XX AAE02435;  
DT 10-AUG-2001 (first entry)  
DE Novel human protein (NHP) #6, sharing similarity with Notch ligand.  
KW Human; neurotrophic; neuroprotective; antidiabetic; cytotostatic; SEL-1; NHP;  
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
KW cholesterol metabolism; coronary artery disease; gene therapy;  
KW ceterbroprotective.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 393  
FT /note= "Encoded by RAA"  
XX  
PN WO200136636-A2.  
XX  
PD 25-MAY-2001.  
XX  
PF 16-NOV-2000; 2000WO-US31373.  
XX  
PR 17-NOV-1999; 99US-0165959.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
DR WPI; 2001-355635/37.  
DR N-PSDB; AAD06379.  
XX  
PT Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX

PS Disclosure; Page 36-37; 39pp; English.  
XX The present sequence is a novel human protein (NHP) which share  
CC structural similarity with animal Notch ligands, particularly SEL-1.  
CC SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 576 AA;  
Query Match 78.2%; Score 2792.5; DB 22; Length 576;  
Best Local Similarity 81.0%; Pred. NO. 1.3e-231;  
Matches 558; Conservative 7; Mismatches 11; Indels 113; Gaps 4;  
QY 1 MKPLSLLEILLIGVTIKTAEHNKROKERNVTTQVSNVEIKQYLSHILEQRTSSNV 60  
Db 1 mkplsllleillilgtvtiktaeehnkrqkernvttqvsneikqylshileqrtssnv 60  
QY 61 INKRENLEKKKNQKIRIKGIONKILRNKNHLQKQAEKNFTDEGDLQFKMGKLVQLQ 120  
Db 61 inkrenllekkknqrirkigiknkdilrnknhlqkqaeKNFTdegdlqfmgkvlqqlq 120  
QY 121 SKSQKQEEAYLLFAKADMGNLKAMEKMAADALLFNGFVQNTAAIQLYESLAKEGSK 180  
Db 121 sksqkqeeayllfakaadmgnlkamekmaadallfngfvgvqntaaqlyleslakegsk 180  
QY 181 AQNALGFLSSYIGMEYDOAKALIIYTTGSGAGNMMSOMILGYRYLSGINVLQNCCEVALS 240  
Db 181 aqnalglfssyigmeYdqakaliYttfgsagnmmsqmilgyrylsginvlqncevals 240  
QY 241 YKKKVDYIADTFEKGSEGVPEKVRTERPENLSSNSETLDDWIIYQYKFLAERGDVQIQ 300  
Db 241 ykkkvadyiadtfeKsegvpekvrtterpenlssnselldwiiyqykflaergdvqilq 300  
QY 301 VSLGOLHLITGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVPQNNATAFK 360  
Db 301 vslgqlhligrkgldqdykalyhflkaakagsanamafigmylegnaavpqnntafk 360  
QY 361 YFSMAASKGNAIGLHGLLPHGKGVPLNYAEALKYQKAEEKGWPDAQFOLGPMYYSYG 420  
Db 361 yfsmaaskgnaiglhlgllyfhgkgvplnyaealkyfqkaeekgwpdaqfqlgfmmyysg 420  
QY 421 SGIKWDYKLAIFYLASOGOPALYIYLAQMYATGTGVVRSCTAVELYKGCVELGHWA 480  
Db 421 sgikwdyklafyylasogoplaIyylakmyatgtgvvrsctavelykgvcelghwa 480  
QY 481 EKFLTAYFAYKDGDDISSLVQYALLAEMGYEVAQNSAFILSKKANILEKEKMPMALL 540  
Db 481 ekfltayfaykdgdiddsslvqyalllaemgyevagqnsafilskkanilekempmall 540  
QY 541 LWNRAAIOGNAPARVIGDYHYGYGTGKDYQTAATHYSIAANKYHNAQAMENLAYMYEH 600  
Db 541 lwnraaioGNAPARVIGDYHYGYGTGKDYQTAATHYSIAANKYHNAQAMENLAYMYEH 600  
QY 601 GLIGTKDIHLARRLYDMAAQTSPDAHIPVLFVAVMKLETHLLRLDILFFNQFTTRWNWLK 660  
Db 601 gligtkdihlarrlydmaaqtspdahipvlfvavmkletthllrldilffnqfttrwnwlk 660  
QY 661 DNTIGPHWDLFVIGLIVPGLLILLNNHHG 689  
Db 661 dntigphwDLFvigliVPGllillnnHHG 689





FT Misc-difference 393 /note= "Encoded by RAA"  
 FT WO200136636-A2.  
 PN 25-MAY-2001.  
 PD 16-NOV-2000; 2000WO-US31373.  
 PF 17-NOV-1999; 99US-0165959.  
 PR (LEXI-) LEXICON GENETICS INC.  
 PA Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
 XX WPI: 2001-355635/37.  
 XX N-PSDB; AAD06378.  
 XX Novel isolated human polynucleotides encoding polypeptides sharing  
 PT sequence similarity with mammalian SEL-1 proteins which are negative  
 PT regulators of Notch family receptors, useful for treating diabetes,  
 PT cancer -  
 XX Claim 6; Page 34-35; 39pp; English.  
 XX The present sequence is a novel human protein (NHP) which share  
 CC structural similarity with animal Notch ligands, particularly SEL-1.  
 CC SEL-1 proteins are negative regulators of Notch family receptors.  
 CC Notch receptors and their associated signalling pathways have been  
 CC associated with development, apoptosis, neuron growth and maintenance.  
 CC Labeled NHP probes can be used to screen a human genomic library which  
 CC is helpful for identifying polymorphisms, determining the genomic  
 CC structure of a given locus/allele and designing diagnostic tests. The  
 CC NHP is also useful in screening techniques for drugs which treats  
 CC symptomatic or phenotypic manifestations of perturbing the normal  
 CC function of NHP in the body. Nucleotide constructs encoding functional  
 CC NHPs, antisense, antisense molecules can be used in gene therapy  
 CC approaches for modulating gene expression such as for preventing or  
 CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
 CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
 CC such as Parkinson's disease, stroke, vascular dementia and conditions  
 CC requiring modulation of fat and cholesterol metabolism such as coronary  
 CC artery disease.  
 XX Sequence 499 AA;  
 SQ

Query Match 67.0%; Score 2392; DB 22; Length 499;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-197;  
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKPLSLLEIILGVTIKTAEEHNRKQKERNVTQVSNIEIKOYLSHILEQRTSSNV 60  
 DB 1 mkplslleililgvtiktkaeehnrkqkernvtqvsneikqylshileqrtsnv 60  
 QY 61 INRENLEKKNQRIKRIQNKDILKRNKHLQKQAEKNFTDEGDLFRMGIKVLQ 120  
 DB 61 inrenllekknqrirkriqngkdlkrnkhlqkqaeaknftdegdlfrmgikvlq 120  
 QY 121 SKSOKKEAYLLFAKADMGNLKMEKADALLFGVGNITATQIYESLAKESGCK 180  
 DB 121 sksqkqkeayllfakadmgnlkamekadallfgnfgvgnitaaqlqieslakegsk 180  
 QY 181 AQNALGFLSSYGIMEDYQAKALIYYTFSAGNNMMSQILGYRYSGLNVLQNCVVALS 240  
 DB 181 aqnalgfllssygmeydqakaliyytfsaggnmmsqilgyrlysglnvlnqncvals 240  
 QY 241 YKKVADYADTPEKSEGVPVEKRVTERPENLSSNSEILDWDIYYKFLAERGQVQIQ 300  
 DB 241 ykkvadyadtpeksegevpvekvriterpenlssnselldwdiyykflaergdvqiq 300  
 QY 301 VSLGQLHLGRKGLDDQYKALHYFLKAAKAGSANAMAFITGKMYLEGNAVPONNATAFK 360  
 DB 301 vslgqlhlgrkglddqykalhyflkaakagsanamafitgkmylegnaavpnnatafk 360

Db 301 vslgqlhlgrkglddqykalhyflkaakagsanamafitgkmylegnaavpnnatafk 360  
 QY 361 YFSMAASKGNAIGLHGILLYFHGKGVPLNYAEALKYFQKAAEKGWPDQOQLGFMYSG 420  
 Db 361 yfsmaaskgnaighlgillyfhgkgvplnyaealkyfkaaekgwpdqqlgfmyysg 420  
 QY 421 SGIWKDYKLAFKYFYLASQSGQPLAIYYLAKMYATGTGVVRSRTAVE 468  
 Db 421 sgiwkdyklafkyfyiasqsgqplaiyylakmyatgtgvvrsrtave 468  
 RESULT 7  
 AAE02433  
 ID AAE02433 standard; Protein; 418 AA.  
 XX AC AAE02433;  
 XX DT 10-AUG-2001 (first entry)  
 XX DE Novel human protein (NHP) #4, sharing similarity with Notch ligand.  
 XX KW Human; nootropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
 KW cholesterol metabolism; coronary artery disease; gene therapy;  
 KW cerebroprotective.  
 OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FH Misc-difference 393  
 FT /note= "Encoded by RAA"  
 FT WO200136636-A2.  
 XX 25-MAY-2001.  
 XX 16-NOV-2000; 2000WO-US31373.  
 XX 17-NOV-1999; 99US-0165959.  
 XX (LEXI-) LEXICON GENETICS INC.  
 XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
 WPI: 2001-355635/37.  
 N-PSDB; AAD06377.  
 Novel isolated human polynucleotides encoding polypeptides sharing  
 sequence similarity with mammalian SEL-1 proteins which are negative  
 regulators of Notch family receptors, useful for treating diabetes,  
 cancer -  
 Claim 5; Page 32-33; 39pp; English.  
 The present sequence is a novel human protein (NHP) which share  
 structural similarity with animal Notch ligands, particularly SEL-1.  
 SEL-1 proteins are negative regulators of Notch family receptors.  
 Notch receptors and their associated signalling pathways have been  
 associated with development, apoptosis, neuron growth and maintenance.  
 Labeled NHP probes can be used to screen a human genomic library which  
 is helpful for identifying polymorphisms, determining the genomic  
 structure of a given locus/allele and designing diagnostic tests. The  
 NHP is also useful in screening techniques for drugs which treats  
 symptomatic or phenotypic manifestations of perturbing the normal  
 function of NHP in the body. Nucleotide constructs encoding functional  
 NHPs, antisense, antisense molecules can be used in gene therapy  
 approaches for modulating gene expression such as for preventing or  
 treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
 insulinomas), blood pressure abnormalities, neurodegenerative diseases  
 such as Parkinson's disease, stroke, vascular dementia and conditions  
 requiring modulation of fat and cholesterol metabolism such as coronary  
 artery disease.

CC requiring modulation of fat and cholesterol metabolism such as coronary artery disease.

XX  
SQ Sequence 418 AA;

Query Match 59.7%; Score 2130; DB 22; Length 418;  
Best Local Similarity 100.0%; Pred. No. 8.3e-175;  
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKPLSLLEILLILGVTIKTKAEHNNKROKERNVTVSVNNEIKOYLSHILEQRTSSNV 60  
DB 1 mkplslleillilgvtiktkaeemkrqernvttqsvsneikqylshilleqrtssnv 60  
QY 61 INKRENLEKKKNRKIRIKGIONKDIILKRNKHLQKQAEKNFTDEGDLFKMGIKVLQ 120  
DB 61 inkrenllekkknrkirkigkndilkrnkhqkqeknftdegdlfkmgikvlq 120  
QY 121 SXSQKQKEAYLLFKAADMGNLKAEMKADALLFGNGVQVNTAAIQIYESLAKEGSK 180  
DB 121 sxsqkqkeayllfkaadmgnlkamekmadalifgnfgvqvnitaaiqlieslakegsck 180  
QY 181 AQNALGFLSSYGIGMEYDQAKALIVYTFGSAGGNMMSOMILGYRYLSGINVLQNCV 240  
DB 181 aqnalgfllssygigmeydqakaliytfgsaggnmmsqmilgyrylsginvlnqncevals 240  
QY 241 YKKVADYIADTFERSEGVPEKVRITERPENLSSNSEILDWDIYQYKFLAERGQVQIQ 300  
DB 241 ykkvadadiadtfeksegvpvekvrilterpenlssnseilddiyykflaergdvqiq 300  
QY 301 VSLQGLHIGRGLDQDYKALHYELFKAAGSANAAMAFIGKMYLEGNAAVPQNNATAFK 360  
DB 301 vslgqlhllgrgldqdykalyfkaagksanaamafigkmylegnaavpqnatafk 360  
QY 361 YFSMAASKGNATGLGGLLYFHGKGVPLNYAEALKYFQKAEKGWPDQFOLGFWY 418  
DB 361 yfsmaaskgnatglggllyfhgkgvplnyaealkyfqkaekgwpdqfqlgfwy 418

RESULT 8  
AAV17750  
ID AAV17750 standard; Protein; 794 AA.  
XX  
AC AAV17750;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Human pancreas-specific tumour suppressor TSA305 protein.  
XX  
KW Human; pancreas; TSA305; pancreatic cancer; carcinoma; diagnosis;  
KW therapy; tumour suppressor.  
XX  
OS Homo sapiens.  
XX  
PN WO9928457-Al.  
XX  
PD 10-JUN-1999.  
XX  
PF 25-NOV-1998; 98WO-JP05306.  
XX  
PR 20-APR-1998; 98JP-0126803.  
PR 28-NOV-1997; 97JP-0343789.  
XX  
PA (SAKA ) OTSUKA PHARM CO LTD.  
XX  
PI Harada Y, Ozaki K;  
XX  
DR WPI; 1999-358128/30.  
DR N-PSDB; AAX76577, AAX76578.  
XX  
PT Tumour suppressor gene TSA305 expressed specifically in pancreas,  
PT useful for diagnosis and treatment of pancreatic cancer  
XX

PS Claim 1; Page 38-41; 54pp; Japanese.  
XX  
CC The present sequence represents the protein from a gene designated  
CC TSA305, which is expressed specifically in pancreas tissue. The gene,  
CC sequences hybridizing with it, its expression product, and antibodies  
CC recognizing the expression product, are useful in the investigation,  
CC diagnosis, prevention and treatment of pancreatic cancer.  
XX  
SQ Sequence 794 AA;

Query Match 46.2%; Score 1650; DB 20; Length 794;  
Best Local Similarity 52.3%; Pred. No. 4.5e-133;  
Matches 313; Conservative 118; Mismatches 158; Indels 10; Gaps 3;

QY 96 QKQAEKNFTDEGDLQFKMGIKVLQOSKQKQKAEYLLFAKAADMGNLKAEMKADALLF 155  
DB 172 eeaakrrmqeaeemyqtgmkilngsnkkskrearyrlylkaasmhktalervsyalif 231  
QY 156 GNFGVQNTAAIQIYESLAKEGSKAQNALGFLSSYGIGMEYDQAKALIYTFGSAGNM 215  
DB 232 gdlpqnigaaremfeklteegspkqtalglfyasglgvnssqakalvytfgalgnl 291  
QY 216 MSQMILGYRYLSGINVLQNCVALLSYKKVADYIADTFEKSEGVPEKVRITERPENLSS 275  
DB 292 iahmvlgvrywagigvlqscsalthyrilvanhvasdisltggsvvqriripdevnpgm 351  
QY 276 NSEILDWDIYQYKFLAERGQVQIOVSLGOLHLIGRKGDLQDYKALHYFLKAAGSANA 335  
DB 352 nsgmleedliqyyqlaekgdvqagvgllhlhgrgvqeqhgrafdyfnlaanagsh 411  
QY 336 AMAFIGKMYLEGNAAVPQNNATAFKYFSAASKNAIGLHGLLGYFHGKGVPLNYAEAL 395  
DB 412 amafigkmysegdivpqsnetaalyfkaadmgnpvgqslgmalylygrgvqvydlal 471  
QY 396 KYFQAAEKGWPDQFOLGFWYSGGIWKDYKLAFFYLASQSGOPLAIYYLAKMYAT 455  
DB 472 kyfqaaekgwvgdqlqigsmyngigvkrdykqalkyfnlasqgghilafynlaqmh 531  
QY 456 GTGVVRSCTAVELYKVCGLHMAEKFALTAYFAYKDGIDISLVQYALLAEMGYEVAOS 515  
DB 532 gtgvmschtavellfnvcergvrserintaynsykdgynaavlylllaeqgyevags 591  
QY 516 NSAFILESKANILEKEMKYPMALLLWNRAAIOGNAFARVKGIDYHYGYGKDYQTAA 575  
DB 592 naafildqreasivgenetyprallhwnraasggytvariklgyhfygfdvdyetaf 651  
QY 576 THYSTAANKYHNAQAMFNLAYWEHGLGITKDIHLARRLYDMAAQTSPDAHLPVLFAYMK 635  
DB 652 lhyrlaseqghsaqamfnlgyhmhexglgkqdihlakrfydmaseasdaqvpvflalck 711  
QY 636 LETTHLL---RDILFFNQFTTRWNWKLKDNLTIGPHWDLF---VIGLIVPGLILLRNHH 688  
DB 712 lgvvyflqyiretnirdmftq---ldmdqlligpewdylmtiallgtviayqrqh 766

RESULT 9  
AAB41251  
ID AAB41251 standard; Protein; 794 AA.  
XX  
AC AAB41251;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORF1015 polypeptide sequence SEQ ID NO:2030.  
XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antiporiatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antitumour; antihypertensive;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;



CC with their binding can be used in a method for treating or preventing  
 CC Alzheimer's disease, diabetes (especially insulin dependent diabetes  
 CC mellitus), cancer (especially pancreatic cancer), stroke, vascular  
 CC dementia, Parkinson's disease, or coronary heart disease. The  
 CC compositions can also be used to treat conditions requiring modulation of  
 CC fat or cholesterol metabolism.

XX Sequence 766 AA;

Query Match 46.1%; Score 1646; DB 20; Length 766;  
 Best Local Similarity 46.5%; Pred. No. 9.4e-133;  
 Matches 334; Conservative 136; Mismatches 204; Indels 44; Gaps 8;

Qy 4 LSLLEILILIGVTIKTI--KAEEHNKQKERNVTVQSVNEIKOYLSHILEQRTSSNVI 61  
 Db 32 Lslilcavllgsaaatddktngdsldskssiptdesvk-----dhdtgkvv 81  
 Qy 62 NKR-----ENLLEKKKQKRIKIGI-----QNKDILKRNKHL----- 95  
 Db 82 agqifvdeeeaevesllqdeedssktgeelsfilesppnssktyeelkrvrkpvltalea 141  
 Qy 96 -OKOAEKNFTDEGDLFKMGTKVLOQSKSQKQKEEAYLLFAKADMGNLKAMEKMADALL 154  
 Db 142 eedaakrrmqdeamelyqagmkilngsnrksqkreayrlylqkaagmnhktalervsyall 201  
 Qy 155 FGNFGVQVNITAAIQIYESLAKEGSCKAQNALGFLSSYGIGMEYDQAKALIVYTFGSAGGN 214  
 Db 202 fgydltqnigaakemfekteegspkggtgilyasglvnsqakalvyttfgalgn 261  
 Qy 215 MMSOMILGYRYLGINVLQNEVALSYKKVADYIADTFEKGVPVKEVRLTERPENLS 274  
 Db 262 liahmilgyrywagilvqscsalthrylvanhvasdisitggsvvqiripdevenpg 321  
 Qy 275 SNSEILDWDIYYKFLAERGQVQIQVSLGOLHILGRKGLDQDYKALHYFLKAAKAGSA 334  
 Db 332 mnsqgmleedliqyvgflaekgdvqvglgqlhghgrgveqnhqrafdyfnlaanagns 381  
 Qy 335 NAMAFIKMYLEGNAVPONNATAFKYFSMAASKGNALGLHGLLGYPHGKGVPLNTAE 394  
 Db 382 hamafikmysegdipqgsnetalhyfkkaadmgnpvgqsglmaylygrgvqynyda 441  
 Qy 395 LKYFQKAEKQWPDQFQGLPMYYSGGIWKDYKLAIFYLASOGSOPLAIIYLA-KMY 453  
 Db 442 lkyfqkaegwgdqqlgsmymyngivkrdykqalkyfnlasggghillafynlaxkmh 501  
 Qy 454 ATGTGVVRSCTAVELYKVCGLHWARKFLTAYFAVKDGDIDSSLYOYLLAEWGEVA 513  
 Db 502 asgtgmrschtavelfknvcegrwserlmtaynsykdedyaaavvqyllaeggyeva 561  
 Qy 514 QSNFAFILESKANILEKERMYPMALILWNRAAIQGNARFARVKIGDYHYGYTKDYOT 573  
 Db 562 qsnafildqreativgenetyprallhwnraasggytvarikigdyhfygfdvdyet 621  
 Qy 574 AATHYSIAANKYHNAQFNLAAYMEHGLGTTKDIHLARRLYDMAQTSPPAHIPVFAV 633  
 Db 622 afihylaseqhsaqamfnigymhekglgikqdihlakrfymaaespdaqvpvflal 681  
 Qy 634 MKLETHLDRILFPNQFTTWNWKLNDNTGPHWDLF---VGLIVPGLILLLRNHH 688  
 Db 682 cklgvvyflvyirean-irdlftqldmqdlglgpedwlylmtciailigtviayqrqh 738

RESULT 11

AAV18095

ID AAV18095 standard; Protein; 788 AA.

XX AAV18095;

AC AAV18095;

DT 10-AUG-1999 (first entry)

XX Full length mouse Sel-1L protein sequence.

DE

XX

XX

XX

XX

XX

XX

KW Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;  
 KW insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;  
 KW vascular dementia; Parkinson's disease; coronary heart disease;  
 KW fat metabolism; cholesterol metabolism.

XX Mus sp.

OS WO9927088-A2.

XX PN 03-JUN-1999.

XX PD 19-NOV-1998; 98WO-CA01058.

XX PF 28-JUL-1998; 98US-0123549.

XX PR 19-NOV-1997; 97US-0066140.

XX XX (MOUN ) MOUNT SINAI HOSPITAL.

XX PA Bernsteins A, Donoviel D;

XX PI WPI; 1999-357833/30.

XX XX N-PSDB; AAX77012.

XX DR New Sel-1L, nucleic acid molecule useful in the treatment of  
 Alzheimer's disease, diabetes and cancer

XX PT Claim 7; Page 70; 77pp; English.

XX PS This sequence is a Sel-1L (Sel-1 like) protein (also previously  
 known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and  
 the host cell containing it can be used to prepare a Sel-1L protein.  
 Compositions containing the Sel-1L proteins, or peptides that interfere  
 with their binding can be used in a method for treating or preventing  
 Alzheimer's disease, diabetes (especially insulin dependent diabetes  
 mellitus), cancer (especially pancreatic cancer), stroke, vascular  
 dementia, Parkinson's disease, or coronary heart disease. The  
 compositions can also be used to treat conditions requiring modulation of  
 fat or cholesterol metabolism.

XX CC Sequence 788 AA;

XX Query Match 46.1%; Score 1646; DB 20; Length 788;  
 Best Local Similarity 46.5%; Pred. No. 9.8e-133;  
 Matches 334; Conservative 136; Mismatches 204; Indels 44; Gaps 8;

Qy 4 LSLLEILILIGVTIKTI--KAEEHNKQKERNVTVQSVNEIKOYLSHILEQRTSSNVI 61  
 Db 32 Lslilcavllgsaaatddktngdsldskssiptdesvk-----dhdtgkvv 81

Qy 62 NKR-----ENLLEKKKQKRIKIGI-----QNKDILKRNKHL----- 95  
 Db 82 agqifvdeeeaevesllqdeedssktgeelsfilesppnssktyeelkrvrkpvltalea 141

Qy 96 -OKOAEKNFTDEGDLFKMGTKVLOQSKSQKQKEEAYLLFAKADMGNLKAMEKMADALL 154  
 Db 142 eedaakrrmqdeamelyqagmkilngsnrksqkreayrlylqkaagmnhktalervsyall 201

Qy 155 FGNFGVQVNITAAIQIYESLAKEGSCKAQNALGFLSSYGIGMEYDQAKALIVYTFGSAGGN 214  
 Db 202 fgydltqnigaakemfekteegspkggtgilyasglvnsqakalvyttfgalgn 261

Qy 215 MMSOMILGYRYLGINVLQNEVALSYKKVADYIADTFEKGVPVKEVRLTERPENLS 274  
 Db 262 liahmilgyrywagilvqscsalthrylvanhvasdisitggsvvqiripdevenpg 321

Qy 275 SNSEILDWDIYYKFLAERGQVQIQVSLGOLHILGRKGLDQDYKALHYFLKAAKAGSA 334  
 Db 332 mnsqgmleedliqyvgflaekgdvqvglgqlhghgrgveqnhqrafdyfnlaanagns 381

Qy 335 NAMAFIKMYLEGNAVPONNATAFKYFSMAASKGNALGLHGLLGYPHGKGVPLNTAE 394  
 Db 382 hamafikmysegdipqgsnetalhyfkkaadmgnpvgqsglmaylygrgvqynyda 441



```
PD 03-JUN-1999.
XX
XX 19-NOV-1998; 98WO-CA01058.
XX
XX 28-JUL-1998; 98US-0123549.
PR 19-NOV-1997; 97US-0086140.
XX
XX (MOUN ) MOUNT SINAI HOSPITAL.
XX
XX Bernstein A, Donoviel D;
PI
XX WPI; 1999-357833/30.
DR N-PSDB; AAX77014.
XX
XX New Sel-1L nucleic acid molecule useful in the treatment of
PT Alzheimer's disease, diabetes and cancer
XX
XX Claim 7; Page 77; 77pp; English.
XX
XX This sequence is a Sel-1L (Sel-1 like) protein (also previously
CC known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and
CC the host cell containing it can be used to prepare a Sel-1L protein.
CC Compositions containing the Sel-1L proteins, or peptides that interfere
CC with their binding can be used in a method for treating or preventing
CC Alzheimer's disease, diabetes (especially insulin dependent diabetes
CC mellitus), cancer (especially pancreatic cancer), stroke, vascular
CC dementia, Parkinson's disease, or coronary heart disease. The
CC compositions can also be used to treat conditions requiring modulation of
CC fat or cholesterol metabolism.
XX
XX Sequence 404 AA;
SQ
Query Match 30.8%; Score 1098; DB 20; Length 404;
Best Local Similarity 54.9%; Pred. No. 5.1e-86;
Matches 208; Conservative 66; Mismatches 95; Indels 10; Gaps 3;
QY 316 QDYKALHYFLKAAKAGSANAMAFICRMYLEGNAAVPPQNNATAFKYFMSAAKGNALGLH 375
DB 2 qnhqrafyfnlaanagnshamafgkmysegdivpqsnetalhyfkkaadmgppvgs 61
QY 376 GLGLYFHCKGVPNVAEALYKFAAEKGPDPADQFQGFMYSSGKWKDYLAKFY 435
DB 62 glgmaylygrgvnydialkyfkaaeqgwdgdlqgsmymyngivkrdykgalkyfn 121
QY 436 LASQSQPLAIYLYLAKMATGTVGVVRSCTAVELYKGVCELGHWAEKFLTAYFAKGDGI 495
DB 122 lasqgghllafnlaqmhastgvmrschtavelfkncvcergrwseerlmtaynsykdgy 181
QY 496 DSSLVQYALLAEMGYEVAQSNFAFILESKKANILEKEMYPMALLWNRAATQGNARFV 555
DB 182 naavqylllaeqgyevagsnaafildqreasivgenetyprallhwnraasqgytvari 241
QY 556 KIGDYHYGYGKDKYQYATHTSYSTAANKYHNAQAMFNLAYMEHGLGTIKDIHLARLY 615
DB 242 kigdhyfygtdvdyetafilylaseqghsaqamfalgymhekglgkqdihaakrfy 301
QY 616 DMAAQTSPDAHPVLFVAVNKLKLETHLL---RDILFFNFTTRWNWKLKLDNTIGPHWDLF- 671
DB 302 dmaaaspaqvpvflaicklgvvyfilyretnrldmftg---ldmdqligpewdlyl 357
QY 672 --VIGLIVPGLILLRNHH 688
DB 358 mtiallaltgtviayrqrgh 376
RESULT 14
AAY95006
ID AAY95006 standard; Protein; 640 AA.
XX
XX AAY95006;
XX
XX 19-JUN-2000 (first entry)
DB
```

Human secreted protein ve13\_1, SEQ ID NO:52.

Human; secreted protein; cancer; tumour; cardiovascular disorder;  
blood disorder; haemophilia; autoimmune disease; diabetes; inflammation;  
infection; fungal; bacterial; viral; HIV; allergy; arthritis;  
neurodegenerative disease; asthma; contraceptive.

Homo sapiens.

WO200011015-A1.

02-MAR-2000.

24-AUG-1999; 99WO-US19351.

24-AUG-1998; 98US-0097638.

24-AUG-1998; 98US-0097659.

09-SEP-1998; 98US-0099618.

28-SEP-1998; 98US-0102092.

25-NOV-1998; 98US-0109978.

23-DEC-1998; 98US-0113645.

23-DEC-1998; 98US-0113646.

23-AUG-1999; 99US-0379246.

(ALPH-) ALPHAGENE INC.

Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;

WPI; 2000-224657/19.

New secreted or transmembrane proteins and polynucleotides encoding  
them, useful for treating neurodegenerative disorders, autoimmune  
diseases and cancer.

Claim 61; Page 312-314; 357pp; English.

The invention relates to 40 human secreted proteins (AAY94981-Y95020),  
and cDNA sequences encoding them (AAA23423-A23462). The secreted  
proteins of the invention include those that are thought to be only  
partially secreted, i.e., transmembrane proteins. The proteins of the  
invention may exhibit one or more activities selected from the following:  
cytokine activity; cell proliferation; differentiation; immune  
modulation; haematopoiesis regulation; tissue growth activity;  
and thrombolytic activity; chemotactic/chemokinetic activity; and tumour  
inhibition activity. The proteins may be administered to patients as  
vaccines, and the nucleotides may be used as part of a gene therapy  
regime. Diseases or conditions that may be treated using the proteins or  
nucleotides of the invention include autoimmune diseases; genetic  
disorders; haemophilia; cardiovascular diseases; cancer; bacterial,  
fungal and viral infections, especially HIV; multiple sclerosis;  
rheumatoid arthritis; pulmonary inflammation; Guillain-Barre syndrome;  
insulin dependent diabetes mellitus; and allergic reactions such as  
asthma and anaemia. They may also be used for treating wounds, burns,  
ulcers, osteoporosis, osteoarthritis, periodontal diseases, Alzheimer's  
disease, Parkinson's disease, Huntington's disease and anyotrophic  
lateral sclerosis (ALS). Proteins with activin/inhibin activity may  
additionally be useful as contraceptives. Nucleic acid sequences of the  
invention may be used in chromosome mapping, and as a source of  
diagnostic primers and probes. The present sequence represents one of the  
40 proteins of the invention.

Sequence 640 AA;

Query Match 12.2%; Score 435; DB 21; Length 640;  
Best Local Similarity 27.1%; Pred. No. 1.2e-28;  
Matches 160; Conservative 101; Mismatches 254; Indels 76; Gaps 20;

QY 100 EKNFTDEGDLF---KMGIKVLQSQSKQKEAYLLFAKAAADGNLKAAMEKMAADLILF 155  
DB 8 ekeldkpslqfallemdlltvpnrqnesvseigkifeke--vkrissid----- 57





Tue Apr 9 08:52:31 2002

us-09-714-882-2.rag

Page 13

```

QY 605 -----TKDIHLARRLYDMA~AQTSPPDAHIPVLFVAVMKLETHLLRDIL 646
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 855 leidstlhnsnnislqelyercwshsneesfpcslawlyhlrllwgail 905

```

Search completed: April 8, 2002, 15:04:59  
Job time: 198 sec





Db 68 RNEKNYQSIWNETDSORHIYEL---LVQSEOFNNSEATYITLQIHLWQYNFPHNM 124  
QY 139 DMCNLKAMEKMDALLFGNFGVQNTAAIQLYESLAKESGCKAQNALGPLSSYGVGMED 198  
Db 125 TIAH-KYLEFNDLTHFTWH-----SALFDLAVMATGCGASGNDQTVIPO-----D 170  
QY 199 QAKALIYYTFSAGGMMQMSQILGYRLSGINVLQNCVALLSYKRVADYIADTFEKS- 257  
Db 171 SAKALIYQRAQLGNLAKAKQVLAYKYSGFVPRNFHSLVLYRDIAQLRKSYSRDEW 230  
QY 258 -----GVPVEKVR-LTERPENLSNSELIDWDIYQYK 289  
Db 231 DIVFPYTESYNRISDFESGLLKGNSVPSSTVRKRTTRPDIGS-----P 276  
QY 290 FLAERGDVOIQVS---LGOHLHGRKGL---DQD-----YKALHVELKAAGA 331  
Db 277 FIAQVNGVOMTLOIEPMGRFAENGNDGNGINGDEDEASERIRIYYAALNDY-KGIYS 335  
QY 332 GSAN-----AMAFIGKMYLEGNAAVPQNNATAF 359  
Db 336 QSRNCERAKNLLLELYKEFQPHVDNLDPLQVYVYVRCLOLLGHMYFTGSGSKPNHMAE 395  
QY 360 KYP--SMAASKNAIGLHG-----LGLLYFHGKGVPLNTAEALKYFKA-----AEKGWD 408  
Db 396 ELITTSLEISR-RAOGPIGRACIDLGLI---NQYITNTISQAISYMKAMKTQANNG--I 449  
QY 409 AOPOLGFMYSGGIWKDYKLA--PKYFYLASOSGLOPLAIYYLAKMYATGTGVVRSCTA 466  
Db 450 VFEQLSKLATS-----FPEKIGDPNLMETAYLNGFIPAIYEFVAVMESGMNSKSVENT 505  
QY 467 VELYKGVCELGH--WAEKELTAYFAYKGDGIDSSIVOVALLAEMGYEVAQNSAFLESK 524  
Db 506 AYLKFTFVDKNEALMAPKLTAFALINDRSEVALWAYSQLAEQGVETQAQVSAAYLMYQL 565  
QY 525 KANILE----KEKMPMALLLNRAAIQGNAPARKYIGDYHYGYGTCKDYQTAATHYSI 580  
Db 566 PYEFEDPPRTDQRTLAISYTYRAFKQGNIDAGVAGDIYF---OMQONYSKAMALYQG 621  
QY 581 AANKYHNAQAMFNLAYMYERHGLGKTDIHLARRLYDMAAQTSPDAHIPVLFAVMKLETH 640  
Db 622 AALKY-SIQAINNLGYMHEHGLGVNRDFHLAKRYIDQVSEHHRVYLASKLSVLKL---H 677  
QY 641 LLRDILFFNQFTRNNLKLNDTIGPHWD 669  
Db 678 LKSWLTWITR--EKVNYWRKPSPLPNED 704

RESULT 2  
US-08-699-103B-8  
; Sequence 8, Application US/08699103B  
; Patent No. 6107462  
; GENERAL INFORMATION:  
; APPLICANT: Rine, Jasper D.  
; APPLICANT: Hampton, Randolph  
; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING  
; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 2200 Sand Hill Road, Suite 100  
; CITY: Menlo Park  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/699,103B  
; FILING DATE: 16-AUG-1996  
; PRIOR APPLICATION DATA:

Db 68 RNEKNYQSIWNETDSORHIYEL---LVQSEOFNNSEATYITLQIHLWQYNFPHNM 124  
QY 139 DMCNLKAMEKMDALLFGNFGVQNTAAIQLYESLAKESGCKAQNALGPLSSYGVGMED 198  
Db 125 TIAH-KYLEFNDLTHFTWH-----SALFDLAVMATGCGASGNDQTVIPO-----D 170  
QY 199 QAKALIYYTFSAGGMMQMSQILGYRLSGINVLQNCVALLSYKRVADYIADTFEKS- 257  
Db 171 SAKALIYQRAQLGNLAKAKQVLAYKYSGFVPRNFHSLVLYRDIAQLRKSYSRDEW 230  
QY 258 -----GVPVEKVR-LTERPENLSNSELIDWDIYQYK 289  
Db 231 DIVFPYTESYNRISDFESGLLKGNSVPSSTVRKRTTRPDIGS-----P 276  
QY 290 FLAERGDVOIQVS---LGOHLHGRKGL---DQD-----YKALHVELKAAGA 331  
Db 277 FIAQVNGVOMTLOIEPMGRFAENGNDGNGINGDEDEASERIRIYYAALNDY-KGIYS 335  
QY 332 GSAN-----AMAFIGKMYLEGNAAVPQNNATAF 359  
Db 336 QSRNCERAKNLLLELYKEFQPHVDNLDPLQVYVYVRCLOLLGHMYFTGSGSKPNHMAE 395  
QY 360 KYP--SMAASKNAIGLHG-----LGLLYFHGKGVPLNTAEALKYFKA-----AEKGWD 408  
Db 396 ELITTSLEISR-RAOGPIGRACIDLGLI---NQYITNTISQAISYMKAMKTQANNG--I 449  
QY 409 AOPOLGFMYSGGIWKDYKLA--PKYFYLASOSGLOPLAIYYLAKMYATGTGVVRSCTA 466  
Db 450 VFEQLSKLATS-----FPEKIGDPNLMETAYLNGFIPAIYEFVAVMESGMNSKSVENT 505  
QY 467 VELYKGVCELGH--WAEKELTAYFAYKGDGIDSSIVOVALLAEMGYEVAQNSAFLESK 524  
Db 506 AYLKFTFVDKNEALMAPKLTAFALINDRSEVALWAYSQLAEQGVETQAQVSAAYLMYQL 565  
QY 525 KANILE----KEKMPMALLLNRAAIQGNAPARKYIGDYHYGYGTCKDYQTAATHYSI 580  
Db 566 PYEFEDPPRTDQRTLAISYTYRAFKQGNIDAGVAGDIYF---OMQONYSKAMALYQG 621  
QY 581 AANKYHNAQAMFNLAYMYERHGLGKTDIHLARRLYDMAAQTSPDAHIPVLFAVMKLETH 640  
Db 622 AALKY-SIQAINNLGYMHEHGLGVNRDFHLAKRYIDQVSEHHRVYLASKLSVLKL---H 677  
QY 641 LLRDILFFNQFTRNNLKLNDTIGPHWD 669  
Db 678 LKSWLTWITR--EKVNYWRKPSPLPNED 704

; APPLICATION NUMBER: 60/002,581  
; FILING DATE: 17-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Grant D.  
; REGISTRATION NUMBER: 31,259  
; REFERENCE/DOCKET NUMBER: 09272/005001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/322-5070  
; TELEFAX: 650/854-0875  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-699-103B-8

Query Match 5.4%; Score 194; DB 3; Length 95;  
Best Local Similarity 60.0%; Pred. No. 2.4e-10;  
Matches 36; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 564 GYGTKDKDTQTAATHYSIAANKYHNAQAMFNLAYMYERHGLGKTDIHLARRLYDMAAQTSP 623  
Db 2 GLDQVDYETAPIHYRLASEQOHSQAAMFNLCYMHKGLGKIKQDHLAKRFRDYMAAVSQP 61

RESULT 3  
US-09-136-574A-44  
; Sequence 44, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, P.O. Box 457  
; CITY: Spring House  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/136,574A  
; FILING DATE: 19-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/932,571  
; FILING DATE: September 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1751 amino acids  
; TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 44:  
US-09-136-574A-44

Query Match 4.0%; Score 144; DB 4; Length 1751;  
Best Local Similarity 19.5%; Pred. No. 0.0009;  
Matches 106; Conservative 78; Mismatches 164; Gaps 25;

QY 140 MGNLAKAMEKADALLFGNFGVONITAAIOLYESLAKESCKAQNALGFLSSYGIGMEYDQ 199  
DB 1 MQEMRAIKR-----VVSITALLVTLSCFPGIMPVKAYAGTYNYGEALQ--- 46  
QY 200 AKALIYYFGSAGGNMMSOMILGYLSGINVLQNCVALSY---YKVVADYIADTFEKS 256  
DB 47 -KTIMEYEF-QMSGKLPKSVWRNRWGDGLD--DGKDVCLDTGWHIDAGDHV----- 95  
QY 257 EGVVPVEKVLTERPENLSNSILWDIYQYKFLAERGDVQIOLVSLGOLHILGRKGLDQ 316  
DB 96 -----KFNLP-----MSYSASMLGWAVYKDAFVKSKOLE-----HILNQIEWAN 136  
QY 317 DYKKAH-----YFLKAAKAGSANAMAFIGKMYLE-----GNAAVPQNNATAFK-----Y 361  
DB 137 DYFVKCHPSKYYIYQ-----VGDPYVDHNFPGPAEVNMQMRPAYKCDLSNPA 184  
QY 362 FSWAASKGNAIGLHGLLPHGKGVPLNYAALYFOKA-----AEKGWPDADFOLGF 415  
DB 185 SSVVETASLAVSVI-----KERNQAAAYLQHAQDLFEFADTRSDAGYTAAT 237  
QY 416 MYTSGGIWKDYKFLAFKPYLASQSQPLAIYYLAKM-----YATGTGVVRS----- 463  
DB 238 GFTSGGFDLGLWAVALYIATDSS-----YLTAEELMSEYANGTWTWCDDVRY 292  
QY 464 -----RTAVELYKGVE--LGHWAER-----FLTAYFA 489  
DB 293 GTLIMAKITGELYKGAVERNLDHTDRTYTPKGMAYLTGWSLRVYATTAFLACVYA 352  
QY 490 YKDGDDSSLOVYALLA--EMGYEVAQNSAFILSKKANILEKEMYPMALLNRAAI 547  
DB 353 DWSGCDSSNKKTYLNFPAKSQIDYALGSTRSVFVG-----GTNYPQHPHNRHNS 404  
QY 548 QGNFARVKIGDYHY-----GYGTTKDYQTAATHY-----SIAANKYHNAQAMFNLAY 596  
DB 405 WANS--MKIPEHRHILYALVGPGSDSYNDITDVFQNEVACD--YNAGIVGALAK 459  
QY 597 MYE 599  
DB 460 MYQ 462

RESULT 4  
US-08-599-171A-30  
; Sequence 30, Application US/08599171A  
; Patent No. 5814473  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, Patrick V.  
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/599,171A  
; FILING DATE: Concurrently  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HERRON, CHARLES J.  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-38  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 592 AMINO ACIDS  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-599-171A-30

Query Match 3.6%; Score 127; DB 2; Length 592;  
Best Local Similarity 21.4%; Pred. No. 0.0055;  
Matches 120; Conservative 76; Mismatches 180; Indels 186; Gaps 28;

QY 67 LLEKKKQKIRIKIGIONKDIKR-----NKN--HLQKQAE----- 100  
DB 43 IVEKKGKIRELVKALWGDKYKAKTGIGHTRWATGKPTDENAHPTDEKGEFAVVHNGI 102  
QY 101 -KNFTDEGDQFKMGIKVLQSKSQKQKEAYLLFAK--AADM--GNLKAMEKMDALLF 155  
DB 103 IENYLEKEELKEGVK-----RSETDEIVIAHLAKNVRGDLLEAVLKVTKLKGAF 158  
QY 156 GNGFQNTAAIOLYESLAKESCKAQNALGFLSSYGIGMEYDQAKALIYYTFSAGNM 215  
DB 159 -----AVITVHE-----PNRL-----IGVK--QGSPLI-VGLGEGENFL 189  
QY 216 MSQMLGYRLSGINVLQNCVALSYKVKVADYIADTFE--KSEGVPEKVLTERPENL 273  
DB 190 ASDIPAILPYTKKIIVLDGE-----IADLTPDVNIYNFEGEPVSKYVN----- 234  
QY 274 SENSELWDIYQYKFLAERGDVQIOLVSLGOLHILGRKGLDQDYKALHYFLKA----- 328  
DB 235 -----ITPDLVS-----AEKG-----GEKHPMLKEIYEQP--KAINDLKGLSTE 274  
QY 329 -----AKGSANAMAFIGKMYLEGNAAPQN--NATAFYKFSMAAS-KGN 370  
DB 275 DAIPFKLDFRRVLIACGTSYHAGFVGKYWIERFAGVPTVIYASEFRYADVPSDKDI 334  
QY 371 ATGLHCLGLLYFHGKGVPLNYAALYFOKAAEKGWPDADFOLGPMYSGSGTWKDYKLA 430  
DB 335 VIGISQSG-----ETADTKFALQSAKEG-----AFTVGLVNVVVGSAIDRE----- 375  
QY 431 FRYFYLASOSQPLAIYYLAKMYATGTGVVRSORTAVELYKGVCELGHWAEKFLTAIFY-- 488  
DB 376 -SDFSLSLTHAGP-----EIGVAATKTFTAQFTA 402  
QY 489 -----AYKGDIDSSLVQYALLAEMGYEVAQS--NSAFILSKKANILEKEMYPMAL 543  
DB 403 LVALSVSRESEENLIR--LLEKVPSSLVQTLNTABEEVKAQVKKKNNMLYLGRYLN 460  
QY 544 RAAIQGNFARVKIGDYHYGY 565  
DB 461 PTALEG-ALKLKEISYIHAEGY 481

RESULT 5  
US-08-646-590B-30  
; Sequence 30, Application US/08646590B  
; Patent No. 5962283

GENERAL INFORMATION:  
APPLICANT: Warren, Patrick V.  
APPLICANT: Swanson, Ronald V.  
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
CURRENT APPLICATION DATA:  
SOFTWARE: FASTSEQ for Windows Version 2.0  
APPLICATION NUMBER: US/08/646,590B  
FILING DATE: 08-May-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/599,171  
FILING DATE: 09-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/01094  
FILING DATE: 21-January-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/017001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 592 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-646-590B-30

Query Match 3.6%; Score 127; DB 2; Length 592;  
Best Local Similarity 21.4%; Pred. No. 0.0055;  
Matches 120; Conservative 76; Mismatches 180; Indels 186; Gaps 28;

QY 67 LLEKKKNQKIRIKGIONKDILKR-----NKN---HLQKQAE----- 100  
DB 43 IVEKKKGKIRELVKALWGKDYKAKTGTGHTRWATHGKPTDENAHPTDEKGEFVAVHNGI 102

QY 101 -KNFTDEGOLFPMGTVKVLQOSKQKQKEAYLLFAK--AADM--GNLKAMEKMDALLF 155  
DB 103 IENYLELKEELKEGVKF-----RSETDTEVIAHLIAKRYGDLLEAVLTKVKKLGAFAP 158

QY 156 GNFQVQNIATAQLYESLAKESCKQAQNALGFLSSVIGIGMEYDQAKALIYYTFGSGAGNM 215  
DB 159 -----AVITVHE-----PNRL-----IGVK-----QGSPLI-VLGEGENFL 189

QY 216 MSOMILGYRLSGINLQNCVALSYKKVADVADTFE--KSEGVPEVKEVRLTERPENL 273  
DB 190 ASDIPAILPYTKIIVLDGE-----IADLPVTNVIYNEFGEFVSKVEM----- 234

QY 274 SSSNEILDNDIYOYKFLAERGVQVLSGLHLIGRGLDQDYKALHYFLKA----- 328  
DB 235 -----ITPDLVS-----AEKG-----GPKHFMKLEIYEQP--KAINDTLKGLFLSTE 274

QY 329 -----AKAGSANAMAFIGKMYLEGNAAPQN--NATAFYFNSAAS-KGN 370  
DB 275 DAIPFKLDFRVLIIACGTSYHAGEVGKWIYERFAGVPTVIYASEFRYADVPVSDKDI 334

QY 371 AIGLHGLGLLYFHGKGVPLNYAEALKYQKAAEKGPDAQFQFQPMYYSGSGIWKDYKLA 430  
DB 335 VIGISQSG-----ETADTKFALQSAKEG-----AFTVGLNVVVSADRE---- 375

QY 431 FKFYLASQSQGPLAIYYLAKMYATGTGVVRSCTRAVELYKVCGLGHWAKEFLTAYP-- 488  
DB 376 -SDFSLSLTHAGP-----EIGVAATKTFTTAQFTA 402

QY 489 ---AYKDGDDSSILVOYALLAEMGYEVAOS-NSAFILESKKANILEKEKMYPMALLWN 543  
DB 403 LYALSVRESEERENLR--LLEKPSLVEQTUNTAEBEVEKVAEKYMKKKNMMLYLGRYLN 460

QY 544 RAAIQCNAAFARVKIGDYHYGY 565  
DB 461 PIALEG-ALKLKEISYIHAEGY 481

RESULT 6  
US-09-069-226-30  
Sequence 30, Application US/09069226  
Patent No. 6013509  
GENERAL INFORMATION:  
APPLICANT: WARREN, Patrick V.  
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/069,226  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/599,171  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: HERRON, CHARLES J.  
REGISTRATION NUMBER: 28,019  
REFERENCE/DOCKET NUMBER: 331400-38  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 592 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-09-069-226-30

Query Match 3.6%; Score 127; DB 3; Length 592;  
Best Local Similarity 21.4%; Pred. No. 0.0055;  
Matches 120; Conservative 76; Mismatches 180; Indels 186; Gaps 28;

QY 67 LLEKKKNQKIRIKGIONKDILKR-----NKN---HLQKQAE----- 100  
DB 43 IVEKKKGKIRELVKALWGKDYKAKTGTGHTRWATHGKPTDENAHPTDEKGEFVAVHNGI 102

QY 101 -KNFTDEGOLFPMGTVKVLQOSKQKQKEAYLLFAK--AADM--GNLKAMEKMDALLF 155  
DB 103 IENYLELKEELKEGVKF-----RSETDTEVIAHLIAKRYGDLLEAVLTKVKKLGAFAP 158





```
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1147 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-260-5

Query Match 3.3%; Score 118.5; DB 3; Length 1147;
Best Local Similarity 19.6%; Pred. No. 0.094;
Matches 134; Conservative 98; Mismatches 246; Indels 207; Gaps 32;

QY 16 VTIKTKAEHNRKQKERNVTTQVSVNEIKOYLSHILEORTSSNVINKRENLLKKNR 75
DB 560 LTTKGLSPQEAANKL-----IKDFLSS--NKELVGKTLNFKAVADAKNTGN 603

QY 76 KIRIKIQNKDILK--RNKNHLQKQAEKNTDEGDQLFKMGIKVLOQSKQKKEEAYLL 133
DB 604 YDEVKKAQ-KDLEKSLRKREHLEKEVEKKLESKSGKNKNMEAK-----AQANSOKDIEFAL 658

QY 134 FAKAADMG-----NLKAMEK-MADALLFGNFGVQNTAAIQLYESLAKEGSKCAQNA 184
DB 659 INKEARNRARAIAQAQLKGIKRELSKLENVNKNLKDQKSPDEFKNGKNKDFSKAEET 718

QY 185 L-----GFLSSYGIGMEYDQAKALIYYTFSAGGNMMSQMLGYRLSGINVLNCEVALS 240
DB 719 LKALKGSVKDLGINPEW-----ISKVENLNAALN 747

QY 241 YK-----KVADIADTFEKGSEGVPEKVELTERPENLSSNSELWDIYQYKFLA 292
DB 748 EFGNGKNKDFSKYQAKSDLENSVKDVIINQ-KVTDKVDNLN-----QAVSVAK 795

QY 293 ERGDV-QIOVSLGQLHLGRKGLDQDYKALHYFLKAAKAGSANAMAFICKMY-----344
DB 796 ATGDFSRVEQALADLNFSKEQLAQ-----QAQKNESLNARK-KSEIYQSVKNGV 844

QY 345 ---LEGNAAPQNNATAFYKFS-----MAASKGNAIGLHGLGLLYFHGKGVPLNYAEALK 396
DB 845 NGTLVGNLSQAEATTLTKNFSDIKKELNAKLGNFNNNNNGL-----KNEPI-YAKVKN 898

QY 397 -----YFOKAAEKGWPDQAFQLGFMYSYSGGIWKDYKLAFFKLYLASQSQPL 444
DB 899 KKAGQAASLEPIYAQVAKK--VNAKIDRLNQITASGLV-----VQQAAGFPL 944

QY 445 ----ATYYLAKMYATGTGVVRSCTAVELYKGYCELGHWAEEKFLTAYFAKDGDDID----496
DB 945 KRHDKVDLSKV-----GLSRN---QELAQKIDNLNQAVSEAKAGFFGNLEQTDIKLKD 995

QY 497 -----SSLVQVALLAEMGYEVAQNSAFILSKKANILEK-----EKMPYMA 538
```

```
Db 996 STKHNPMLWVESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINKATGMLTKNPEW 1055
QY 539 LLLNRAAIOGNAFARVKIGDYHYGYGPK--KDYQTAATHYIAANKYHNA-----588
Db 1056 LKLVDNKIVAHNV-GSVPLSEYDKIGFNOKNMKDYSDSKF-----STKLNNAVKTNSGF 1110
QY 589 ----QAMFNLAYMY-----EHL 602
Db 1111 TQFLTNAFSTASYICLARENAEHGI 1135

RESULT 9
US-08-471-491-5
; Sequence 5, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471,491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-08-471-491-5

Query Match 3.3%; Score 118.5; DB 3; Length 1147;
Best Local Similarity 19.6%; Pred. No. 0.094;
Matches 134; Conservative 98; Mismatches 246; Indels 207; Gaps 32;

QY 16 VTIKTKAEHNRKQKERNVTTQVSVNEIKOYLSHILEORTSSNVINKRENLLKKNR 75
DB 560 LTTKGLSPQEAANKL-----IKDFLSS--NKELVGKTLNFKAVADAKNTGN 603

QY 76 KIRIKIQNKDILK--RNKNHLQKQAEKNTDEGDQLFKMGIKVLOQSKQKKEEAYLL 133
DB 604 YDEVKKAQ-KDLEKSLRKREHLEKEVEKKLESKSGKNKNMEAK-----AQANSOKDIEFAL 658

QY 134 FAKAADMG-----NLKAMEK-MADALLFGNFGVQNTAAIQLYESLAKEGSKCAQNA 184
DB 659 INKEARNRARAIAQAQLKGIKRELSKLENVNKNLKDQKSPDEFKNGKNKDFSKAEET 718

QY 185 L-----GFLSSYGIGMEYDQAKALIYYTFSAGGNMMSQMLGYRLSGINVLNCEVALS 240
DB 719 LKALKGSVKDLGINPEW-----ISKVENLNAALN 747

QY 241 YK-----KVADIADTFEKGSEGVPEKVELTERPENLSSNSELWDIYQYKFLA 292
DB 748 EFGNGKNKDFSKYQAKSDLENSVKDVIINQ-KVTDKVDNLN-----QAVSVAK 795

QY 293 ERGDV-QIOVSLGQLHLGRKGLDQDYKALHYFLKAAKAGSANAMAFICKMY-----344
DB 796 ATGDFSRVEQALADLNFSKEQLAQ-----QAQKNESLNARK-KSEIYQSVKNGV 844

QY 345 ---LEGNAAPQNNATAFYKFS-----MAASKGNAIGLHGLGLLYFHGKGVPLNYAEALK 396
DB 845 NGTLVGNLSQAEATTLTKNFSDIKKELNAKLGNFNNNNNGL-----KNEPI-YAKVKN 898

QY 397 -----YFOKAAEKGWPDQAFQLGFMYSYSGGIWKDYKLAFFKLYLASQSQPL 444
DB 899 KKAGQAASLEPIYAQVAKK--VNAKIDRLNQITASGLV-----VQQAAGFPL 944

QY 445 ----ATYYLAKMYATGTGVVRSCTAVELYKGYCELGHWAEEKFLTAYFAKDGDDID----496
DB 945 KRHDKVDLSKV-----GLSRN---QELAQKIDNLNQAVSEAKAGFFGNLEQTDIKLKD 995

QY 497 -----SSLVQVALLAEMGYEVAQNSAFILSKKANILEK-----EKMPYMA 538
```

```
Db 945 KRHDVDDLSKV-----GLSRN-----QELAQKIDNLNOAVSEAKAGFFGNLEQOTIDKDKD 995
Qy 497 -----SSLVQYALLAEMGYEVAOSNSAFILSKKANILEK-----EKMPMA 538
Db 996 STKHNPMLNWLWESAKKVPASLKDNTATNSHIRINSIKNGAINEKATGMTOKNPEW 1055
Qy 539 LLLWNRAAIOGNAPARVKIGDYHYGYGK--KYQTAATHYSIAANKYHNA-----588
Db 1056 LKLVNDKIVAHNV-GSVPLSEYDKIGFNQKNMKDYSDFKE-----STKLNNAVKDTNSGF 1110
Qy 589 ----QAMENLAYMY-----EHGL 602
Db 1111 TOFLTNASTASYCLARENAEHGI 1135

RESULT 10
US-08-466-662-5
; Sequence 5, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466.662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-466-662-5

Query Match 3.3%; Score 118.5; DB 4; Length 1147;
Best Local Similarity 19.6%; Pred. No. 0.094;
Matches 134; Conservative 98; Mismatches 246; Indels 207; Gaps 32;

Qy 16 VTIKTIKAEHNKROKERNVTQVSVNEIKOYLSHILEORTSSNVINKRENLLKKKNQR 75
Db 560 LTTKGLSPQEAANKL-----IKDFLSS--NKELVGKTLNFKAVADAKNTGN 603
Qy 76 KIRIKGIONKDLK--RNKNHLQKQAEKNFTDEGDLQFKMGIKVLQOQSKOKKEEAYLL 133
Db 604 YDEVKQAO-KDLEKSLRKRHLEKEVERKLESKSGNKNKMEAK-----AQANSOKDEIFAL 658
Qy 134 FAKAADMG-----NLKAMEK-MADALLFGNFGVQNTAAIQLYESLAKEGSCAKQNA 184
Db 659 INKEANRARAAYAOQLKGLKRELSKLENVNNKLAOFKDSFDFKNGKNKDFSAKET 718
Qy 185 L----GFLSSYGIGMEYDQAKALIYYTFGAGNMMSOMILGYRYLSGINVLQNCVALS 240
Db 719 LKALKGSVKDLGINPEW-----ISKVENLNAALN 747
Qy 241 YK-----KVADYIADTEKSGVPEKVRVLTTERPENLSSNSEILDWDIYQYKFLA 292
Db 748 EFKNGKNKDFSKVTQAKDSLENSKDVITIQ-KVPDKVDNLN-----QAVSVAK 795
Qy 293 ERGDV-QIOVSLGQLHLGRGLDQDYKALHYFLKAAGKAGSANAMAFIGKMY-----344
Db 796 ATGDSRVEQALADLNKFSKEQLAO-----QAKNESLNARK-KSEIYQSVKNGV 844
Qy 345 ----LEGNAVPONNATAFKYFS-----MAASKGNAIGLHGLLYFHGKGVPLNYAEALK 396
Db 845 NGTLVGNLSQAEATVLSKNFSDIKKELNAKLGNNNNNGL-----KNEPI-YAKYNK 898
Qy 397 -----YFOKAAEKGWPDQAQFQLGFMYYSGSGIWKDYKLAIFYLASQSGQPL 444
```

```
Db 899 KKAQOASLEEPIYAQVAKK--VNAKIDRLNQLASGLGV-----VGQAAGFPL 944
Qy 445 ---AIYVLAKMYATGTGVVRSCTRAVELYKGVCELGHWAEKFLTAYFAYKDDID----496
Db 945 KRHDVDDLSKV-----GLSRN-----QELAQKIDNLNOAVSEAKAGFFGNLEQOTIDKDKD 995
Qy 497 -----SSLVQYALLAEMGYEVAOSNSAFILSKKANILEK-----EKMPMA 538
Db 996 STKHNPMLNWLWESAKKVPASLKDNTATNSHIRINSIKNGAINEKATGMTOKNPEW 1055
Qy 539 LLLWNRAAIOGNAPARVKIGDYHYGYGK--KYQTAATHYSIAANKYHNA-----588
Db 1056 LKLVNDKIVAHNV-GSVPLSEYDKIGFNQKNMKDYSDFKE-----STKLNNAVKDTNSGF 1110
Qy 589 ----QAMENLAYMY-----EHGL 602
Db 1111 TOFLTNASTASYCLARENAEHGI 1135

RESULT 11
US-08-477-451-2
; Sequence 2, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3289 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-451-2

Query Match 3.3%; Score 118.5; DB 2; Length 3289;
Best Local Similarity 19.6%; Pred. No. 0.52;
Matches 134; Conservative 98; Mismatches 246; Indels 207; Gaps 32;

Qy 16 VTIKTIKAEHNKROKERNVTQVSVNEIKOYLSHILEORTSSNVINKRENLLKKKNQR 75
Db 2091 LTTKGLSPQEAANKL-----IKDFLSS--NKELVGKTLNFKAVADAKNTGN 2134
Qy 76 KIRIKGIONKDLK--RNKNHLQKQAEKNFTDEGDLQFKMGIKVLQOQSKOKKEEAYLL 133
Db 2135 YDEVKQAO-KDLEKSLRKRHLEKEVERKLESKSGNKNKMEAK-----AQANSOKDEIFAL 2189
```



CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/059,032  
FILING DATE: 19930507  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William E.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 15270-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 689 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-059-032-5

Query Match 3.2%; Score 113; DB 1; Length 689;  
Best Local Similarity 18.6%; Pred. No. 0.13;  
Matches 95; Conservative 86; Mismatches 201; Indels 128; Gaps 19;

QY 32 ERNVTQVSNKQYLHLEORTSSNVINKR-----ENLEKKKNQKIRIKIQNK 85  
DB 72 EYTYTVORNILDPQHVSPSKDIRTASTADKKLSEFVEMSMREDVYQRIWVLEKQVK 131  
QY 86 DIL-----KRNKNHLOKQAEKNFTDEGDLFRMGIKVLOQSKSQKQKEA 130  
DB 132 DLSRPEAARYLERLIKLGRRNGHLHPRETQENI-----KRIKKLSLCLIDFNKNLNEDT 186  
QY 131 YLLFAKAADGN-----LKAMEKMDALLFGNFGVQNIITAAIOLYESLAKE----- 176  
DB 187 TELPFTLQELGGLPEDFLNSLEKMD-----GKLKVTLYKPHYFPLKCKHVPETRRK 239  
QY 177 -----GSKCAQNALGFLSSYGIGMEYDOAKALIYTFGSAGNMMSQMLGYRYLSGIN 230  
DB 240 VEEAFNCRKEENCA--ILKELVTLRAQKSRLLGFTHAD-----YYLEMN 283  
QY 231 VLQNCVALLSYKKVADYIADTFEKSEGVPEKVRILTERPENLSSNSEILDNDIYQY-- 288  
DB 284 MAKTSOTVATFDELAKLPLGEQERAVILELKRAECERRGLPFDGRIRAWDM-RYYMN 342  
QY 289 -----RFLAERGQVQIOLVSLGQL-----HLIGRGLDQDYKALHYFLKAAKAGS 333  
DB 343 QVEETRYCYVDQNLKKEYFPQV-VTHGLLGIYQELLGLAFHHEEGASAHNEDVRLYTARD 401  
QY 334 ANMAFIGNKYLE-----GNAA-----VPONNATATFYFMSAAS-----KGNATG 373  
DB 402 AASGEVVGKRYLDLYPREGKYGHAACFGLQPGCLRGDSGRQIAAAMVANFTKPTADAPS 461  
QY 374 L--HGGLLYFHCKGVPYLAWEALKYFQKAAEKGWDAQFOLGFMWYSSGIGWIKYKFLAF 431  
DB 462 LQHDVEITYFH-----BFGVHMQLCSQAE-----FAMFSGTHVERDEVEAP 504  
QY 432 KYFLASQSQGLAIYYLAKMVTATGTGVVR 461  
DB 505 SQM-LENWVWEQEPFLRMSRHYTGSAVPR 533

RESULT 14

PCT-US91-07290-5  
; Sequence 5, Application PC/TUS9107290  
; GENERAL INFORMATION:  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Seubert, Peter A.  
; APPLICANT: Dovey, Harry F.  
; APPLICANT: McConlogue, Lisa C.  
; APPLICANT: Little, Sheila P.  
; APPLICANT: Johnstone, Edward M.  
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Athena Neurosciences, Inc.  
; STREET: 800F Gateway Blvd.  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/07290  
; FILING DATE: 19911004  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Feix  
; REGISTRATION NUMBER: 31547  
; REFERENCE/DOCKET NUMBER: 17796-002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 877-0900  
; TELEFAX: (415) 877-8370  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 689 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; PCT-US91-07290-5

Query Match 3.2%; Score 113; DB 5; Length 689;  
Best Local Similarity 18.6%; Pred. No. 0.13;  
Matches 95; Conservative 86; Mismatches 201; Indels 128; Gaps 19;

QY 32 ERNVTQVSNKQYLHLEORTSSNVINKR-----ENLEKKKNQKIRIKIQNK 85  
DB 72 EYTYTVORNILDPQHVSPSKDIRTASTADKKLSEFVEMSMREDVYQRIWVLEKQVK 131  
QY 86 DIL-----KRNKNHLOKQAEKNFTDEGDLFRMGIKVLOQSKSQKQKEA 130  
DB 132 DLSRPEAARYLERLIKLGRRNGHLHPRETQENI-----KRIKKLSLCLIDFNKNLNEDT 186  
QY 131 YLLFAKAADGN-----LKAMEKMDALLFGNFGVQNIITAAIOLYESLAKE----- 176  
DB 187 TELPFTLQELGGLPEDFLNSLEKMD-----GKLKVTLYKPHYFPLKCKHVPETRRK 239  
QY 177 -----GSKCAQNALGFLSSYGIGMEYDOAKALIYTFGSAGNMMSQMLGYRYLSGIN 230  
DB 240 VEEAFNCRKEENCA--ILKELVTLRAQKSRLLGFTHAD-----YYLEMN 283  
QY 231 VLQNCVALLSYKKVADYIADTFEKSEGVPEKVRILTERPENLSSNSEILDNDIYQY-- 288  
DB 284 MAKTSOTVATFDELAKLPLGEQERAVILELKRAECERRGLPFDGRIRAWDM-RYYMN 342  
QY 289 -----RFLAERGQVQIOLVSLGQL-----HLIGRGLDQDYKALHYFLKAAKAGS 333  
DB 343 QVEETRYCYVDQNLKKEYFPQV-VTHGLLGIYQELLGLAFHHEEGASAHNEDVRLYTARD 401









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2002, 15:05:26 ; Search time 13.36 Seconds  
(without alignments)  
1890.873 Million cell updates/sec

Title: US-09-714-882-2

Perfect score: 3570

Sequence: 1 MKPLSLIETILILGVTIKT.....LFVIGLIVPGILLLRNHHG 689

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	1650	46.2	794	1	SEIL_HUMAN	Q9ubv2	homo sapien
2	1635	45.8	790	1	SEIL_MOUSE	Q922g6	mus musculus
3	314	8.8	327	1	YBEQ_ECOLI	P77234	escherichia
4	219	6.1	696	1	SRT5_YEAST	P34226	saccharomyc
5	180	5.0	932	1	YAIA_SCHPO	Q09897	schizosacch
6	179	5.0	515	1	Y141_HUMAN	Q14154	homo sapien
7	168.5	4.7	250	1	HCPA_HELPY	Q25001	helicobacte
8	161.5	4.5	250	1	HCPA_HELPY	Q25001	helicobacte
9	158.5	4.4	184	1	YBET_ECOLI	P77296	escherichia
10	141.5	4.0	567	1	Y228_METJA	Q58823	methanococc
11	137	3.8	512	1	YES6_YEAST	P39000	saccharomyc
12	136	3.8	2869	1	RBPI_PLAYB	Q00798	plasmodium
13	133	3.7	267	1	EXOR_RHILV	Q52822	rhizobium 1
14	132.5	3.7	165	1	Y225_HAEIN	P44277	haemophilus
15	130.5	3.7	334	1	Y798_METJA	Q58208	methanococc
16	126.5	3.5	1150	1	YKDI_CAEEL	Q03560	caenorhabdi
17	125.5	3.5	268	1	EXOR_RHIME	Q52926	rhizobium m
18	125	3.5	1694	1	CLXR_DICTDI	P25870	dictyosteli
19	124	3.5	591	1	GLMS_AQUAE	O66648	a glucosami
20	122.5	3.4	3911	1	AKA9_HUMAN	Q95996	h a kinase
21	119.5	3.3	723	1	DPF3_DROME	Q9vnr8	drosophila
22	118.5	3.3	724	1	HMMR_HUMAN	O75330	homo sapien
23	118.5	3.3	1147	1	CGAI_HELPY	P80200	helicobacte
24	116	3.2	716	1	YOGF_BAGSU	P54488	bacillus su
25	114	3.2	864	1	Y3M_YEAST	P48237	saccharomyc
26	114	3.2	945	1	ANPE_MOUSE	P16406	mus musculus
27	114	3.2	1235	1	DPOL_PYRHO	Q09610	pyrococcus
28	114	3.2	1738	1	YCFI_EPIVI	Q00383	epifagus vi
29	113	3.2	688	1	MEPD_HUMAN	P52888	homo sapien
30	112.5	3.2	514	1	TRPE_EUCRM	Q44689	buchnera ap
31	112.5	3.2	1286	1	RPOL_VARV	P33053	variola vir
32	112.5	3.2	1699	1	DPOL_THEG8	Q9h84	thermococcu
33	112	3.1	1928	1	MYSL_YEAST	P08964	saccharomyc

#### RESULT 1

ID	SEIL_HUMAN	STANDARD;	PRT;	794 AA.
AC	Q9UBV2; Q9P1T9; Q9UHK7;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	SEL-1 HOMOLOG PRECURSOR (SUPPRESSOR OF LIN-12-LIKE PROTEIN) (SEL-1L).			
GN	SEL1L OR TSA305.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pancreas;			
RX	MEDLINE=99426085; PubMed=10496078;			
RA	Harada Y., Ozaki K., Suzuki M., Fujiwara T., Takahashi E.-I.,			
RA	Nakamura Y., Tanigami A.;			
RT	"Complete cDNA sequence and genomic organization of a human			
RT	pancreas-specific gene homologous to Caenorhabditis elegans sel-1.";			
RL	J. Hum. Genet. 44:330-336(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pancreas;			
RX	MEDLINE=20208354; PubMed=10746565;			
RA	Biunno I., Bernard L., Dear P., Cattaneo M., Volorio S., Zannini L.,			
RA	Bankier A., Zollo M.;			
RT	"SEL1L, the human homolog of C. elegans sel-1: refined physical			
RT	mapping, gene structure and identification of polymorphic markers.";			
RL	Hum. Genet. 106:227-235(2000).			
CC	-!- FUNCTION: MAY PLAY A ROLE IN NOTCH SIGNALING (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS.			
CC	-!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-II DOMAIN.			
CC	This SWISS-PROT entry is copyright.. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AB024763; BAA89204.1;			
DR	EMBL; AB024747; BAA89204.1; JOINED.			
DR	EMBL; AB024748; BAA89204.1; JOINED.			
DR	EMBL; AB024749; BAA89204.1; JOINED.			
DR	EMBL; AB024750; BAA89204.1; JOINED.			
DR	EMBL; AB024751; BAA89204.1; JOINED.			
DR	EMBL; AB024752; BAA89204.1; JOINED.			
DR	EMBL; AB024753; BAA89204.1; JOINED.			
DR	EMBL; AB024754; BAA89204.1; JOINED.			
DR	EMBL; AB024755; BAA89204.1; JOINED.			
DR	EMBL; AB024756; BAA89204.1; JOINED.			
DR	EMBL; AB024757; BAA89204.1; JOINED.			

#### ALIGNMENTS

DR EMBL; AB024758; BAA89204.1; JOINED.  
 DR EMBL; AB024759; BAA89204.1; JOINED.  
 DR EMBL; AB024760; BAA89204.1; JOINED.  
 DR EMBL; AB024761; BAA89204.1; JOINED.  
 DR EMBL; AB024762; BAA89204.1; JOINED.  
 DR EMBL; AB020335; BAA87904.1; -.  
 DR EMBL; AF052059; AAF29413.1; -.  
 DR EMBL; AF157516; AAF24176.1; -.  
 DR HSP; P02751; 2FN2.  
 DR MIM; 602329; -.  
 DR InterPro; IPR000562; FN\_Type\_II.  
 DR Pfam; PF00040; fn2; 1.  
 DR PRINTS; PR00013; FNTYPEII.  
 DR ProDom; PD000995; FN\_Type\_II; 1.  
 DR SMART; SM00059; FN2; 1.  
 DR PROSITE; PS00023; FIBRONECTIN\_2; 1.  
 KW Signal; Transmembrane; Glycoprotein.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 794 SEL-1 HOMOLOG.  
 FT TRANSMEM 278 298 POTENTIAL.  
 FT TRANSMEM 701 721 POTENTIAL.  
 FT TRANSMEM 739 759 POTENTIAL.  
 FT DOMAIN 127 168 FIBRONECTIN TYPE-II.  
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 608 608 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 186 186 M -> V (IN REF. 2).  
 FT SEQUENCE 794 AA; 88754 MW; 323EB03DC7485459 CRC64;  
 Query Match 46.2%; Score 1650; DB 1; Length 794;  
 Best Local Similarity 52.3%; Pred. No. 4.6e-100;  
 Matches 313; Conservative 118; Mismatches 158; Indels 10; Gaps 3;  
 QY 96 QKQAEKNTDEGDLQFKMGIKVLQSKSQKQEEAYLLFAKADMGNLKAMEKMDALLF 155  
 DB 172 EBAKRRQOEAEEMMYQTKTILGNSKRSQREAYRYLQKAASNMHTKALERSVALLF 231  
 QY 156 GNFGVQNTAAATOLVESLAKESCAQNALGFLSSVIGMEYDQAKALIYTFGSAGNM 215  
 DB 232 GYLQNTQAAEMFEKLTSESPKQTFALFYASGLGVNSQAKALIYTFGALGML 291  
 QY 216 MSQMLGYRLSGINQCNVALSYKKVADYIADTFEKSQGVPEVKVRLTERPENLSS 275  
 DB 292 IAHVLYGYRWAGIGVQSCSALTHYRLVANHVASDISLTGGSVYQRLRDEVENPCM 351  
 QY 276 NSEILDWDIYQYKFLAERGDVQIQVSLGQLHLGRKLDQDYKALHYFLKAAKAGSAN 335  
 DB 352 NSGMLDEDLIQYQFLAERGDVQIQVSLGQLHLGRGVEQHQRAFDFYFLAANAGNSH 411  
 QY 336 ANAFIGKMYLEGNAAPQNNATAFKYESMAASKGNAIGLGLGLFYHCKGVPLNVAEAL 395  
 DB 412 ANAFIGKMTSEGSDDIVPSNETALHYFKKAADMGPVGSGGLGMAYLYRGVQVNYDUAL 471  
 QY 396 KYFOAAEKGWDAQFLQGLFMYTSGSINKDKYKAFKPYFLASQSQQPIAIYLAQMVA 455  
 DB 472 KYFOAAEKGWDAQFLQGLFMYTSGSINKDKYKAFKPYFLASQSQQPIAIYLAQMVA 531  
 QY 456 GTGVRSCTAVELYKGVCELGHWAEKFTATYFAYKDGDISLSVOYALLAEMGEVVAQS 515  
 DB 532 GTGVRSCTAVELYKGVCELGHWAEKFTATYFAYKDGDISLSVOYALLAEMGEVVAQS 591  
 QY 516 NSAFILSKANILKEKMYPNALLWNRAATQGNARVKIGDYHYGYGPKDYQRTAA 575  
 DB 592 NAAFILDQREASIVGENETYPALHLWNRAASQGTVAIKLGDHFFYGFQVDYETAF 651  
 QY 576 THYSTAANKYHNAQAMFNLAYMYEHLGITKDIHLARLYDMAAQTSPDAHIPVLFAVMK 635  
 DB 652 IHYRLASEQHQSAQAMFNLYGMHEKGLGKIQDIHLAKRFYDMAEASPDQVPELALCK 711  
 QY 636 LETTHLL---RDILFFNFTTRNNMLKLDNTIGPHWDLF---VIGLIVPGLILLRNHH 688

DB 712 LCVWVFLQVIRETNRDMETQ-----LDMQLLCPWDLYLMTIALLLTGTVIAYRQRQH 766  
 RESULT 2  
 SEIL\_MOUSE  
 ID SEIL\_MOUSE STANDARD; PRT; 790 AA.  
 AC Q942G6; Q9DBD8;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SEL-1 HOMOLOG PRECURSOR (SUPPRESSOR OF LIN-12-LIKE PROTEIN) (SEL-1L).  
 GN SEL1 OR SEL1H  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=99077704; PubMed=9858735;  
 RA Donoviel D.B., Donoviel M.S., Fan E., Hadjantonakis A.-K.,  
 RA Bernstein A.;  
 RA "Cloning and characterization of Sel-1L, a murine homolog of the C.  
 elegans Sel-1 gene.";  
 RL Mech. Dev. 78:203-207(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX STRAIN=C57BL/6J; Tissue=Liver;  
 MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saio T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saio R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Offelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzairelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN NOTCH SIGNALING (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-II DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; AF063095; BAB05210.1; -.  
 DR EMBL; AK005023; BAB23750.1; -.  
 DR MGD; MGI:1329016; Sellh.  
 DR ProDom; PD000995; FN\_Type\_II; 1.  
 KW Signal; Transmembrane; Glycoprotein; Alternative splicing.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 790 SEL-1 HOMOLOG.  
 FT TRANSMEM 274 294 POTENTIAL.  
 FT TRANSMEM 597 717 POTENTIAL.  
 FT TRANSMEM 735 755 POTENTIAL.

RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12 / MG1655;
RX	MEDLINE=97426617; PubMed=9278503;

Query Match	8.8%;	Score 314;	DB 1;	Length 327;
Best Local Similarity	29.8%;	Pred. No. 1.5e-13;		
Matches 103;	Conservative 57;	Mismatches 134;	Indels 52;	Gaps 10;

  

Qy	271	ENLSNSPILDWIDYQYYKFLAERGDYQIOVSLGOLHLIGRKGLDQ--DYTKALHYFLKAA	329
Db	11	DNLSID-EIIE-----RAEGDCEAQIIVG-FYNRDSALDSPDEKAFYWLKAA	59
Qy	330	KAGSANAMAF1GKMYLEGNAAPVONNATAFYF5MAASKGNA1GHLGGLLYFHFGKGVPL	389
Db	60	EQGHCEAQYSLGQYTEDKSRHKDNEQAIF-WLKKAALQGHFTFASNALG--WTLDRGEAP	116
Qy	330	NYAEALKYFQAAEBKGWPDQAQFQLGFMYYSSG1WKDYKFLAKFYFYLASGSGQPLATYYL	449
Db	117	NYKEAVVYQIAAESGMSYQAONN1GWMYRNGGVAKDYALAFEFWYKQAALQGHGSDAQNNL	176
Qy	450	AKMVA1TGTVGVRSCTAVELYKGVCELGHWAEEKFLTAFAYKCGDIDSSLVQVALLAEMG	509
Db	177	ADLVGDKGVQVN-----KTLA1FW-----YUKSAQQG	204
Qy	510	YEVAQNSAF1LESKKANILEKMYPMALLMNRRA1TQGNAFARVKIGDYHYHYGYGTTK	569
Db	205	NRHAQFQ1AWDYNAGEG---VDQDYKQAMWY1LKAAQAQGSVGAVN1GYMYKHGQGVK	260
Qy	570	DYQTAATHYSTAANKYHNQAAMFN1AYWEHGLGITKD1HILARLY	615
Db	261	DYQAAEFWFTKAA--ECNDATAYN1LA1MYHYGEGRPVDLRQALDY	305

```

RESULT 4
SKT5_YEAST
ID SKT5_YEAST STANDARD; PRT; 696 AA.
AC P34226; Q02215;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE SKT5 PROTEIN.
GN SKT5 OR CAL2 OR CHS4 OR YBL061C OR YBL0506 OR YBL0519.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94205266; PubMed=8154187;
RA Scherens B., el Bakkoury M., Vierendeels F., Dubois E., Messenguy F.;
RT "Sequencing and functional analysis of a 32,560 bp segment on the
RT left arm of yeast chromosome II. Identification of 26 open reading
RT frames, including the KIP1 and SEC17 genes.";
RL Yeast 9:1355-1371(1993).
RN [2]
RP SEQUENCE OF 1-678 FROM N.A.
RC STRAIN=KR13;
RX MEDLINE=93379362; PubMed=7764021;
RA Kawamoto S., Sasaki T., Itahashi S., Hatsuyama Y., Ohno T.;
RT "A mutant allele skt5 affecting protoplast regeneration and killer
RT toxin resistance has double mutations in its wild-type structural
RT gene in Saccharomyces cerevisiae.";
RL Biosci. Biotechnol. Biochem. 57:1391-1393(1993).
RN [3]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RA Kawamoto S., Nomura M., Ohno T.;
RT "Cloning and characterization of SKT5, a Saccharomyces cerevisiae
RT gene that affects protoplast regeneration and resistance to killer
RT toxin of Kluyveromyces fragilis.";
RL J. Ferment. Bioeng. 74:199-208(1992).
RN [4]
RP SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=97461567; PubMed=9314530;
RA Demarini D.J., Adams A.E., Fares H., De Virgilio C., Valle G.,
RA Chuang J.S., Pringle J.R.;
RT "A septin-based hierarchy of proteins required for localized
RT deposition of chitin in the saccharomyces cerevisiae cell wall.";
RL J. Cell Biol. 139:75-93(1997).
CC -!- FUNCTION: POSSIBLE ROLE IN PROTOPLAST REGENERATION AND KILLER
CC TOXIN OF K.LACTIS (PGKL) RESISTANCE.
CC -!- SUBUNIT: MAY INTERACT WITH CHS3 AND SEEMS TO BE AN ADAPTOR (ALONG
CC WITH BNI4) TO LINK CHS3 TO SEPTINS.
CC -!- SIMILARITY: SOME, TO S.POMBE SPAC24B11.10C AND TO YEAST YER096W.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-15 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 23261; CA80786.1; -
CC EMBL; 235823; CA84882.1; -
CC EMBL; S65415; AAC60564.1; ALT_INIT.
CC PIR; S39827; S39827.
CC PIR; S37328; S37328.
CC SGD; S0000157; SKT5.
CC Prenylation; Lipoprotein.
KW LIPID
FT 693 693 FARNESYL (BY SIMILARITY).
FT CONFLICT 350 350 G -> E (IN REF. 2).
FT CONFLICT 643 643 Q -> T (IN REF. 2).
FT SEQUENCE 696 AA; 77066 MW; D50DEB25E175D165 CRC64;

Query Match 6.1%; Score 219; DB 1; Length 696;
Best Local Similarity 21.6%; Pred. No. 6.7e-07;
Matches 98; Conservative 83; Mismatches 192; Indels 80; Gaps 14;

QY 31 KERNVTQVSNEIKOY-----LSH--ILEQRTSSNVINKRENLEKKKQNRKIRIKGIO 83
DB 129 RQTNSTSLTKKEIKKRTRSDLSHMYLLNGSSDTQTATNESVADLSHQMSRYLGGKN 188
QY 84 NKDILKRNKN-HLQKQAEKNFTDEGDLQPKMGKIKVLOQS----- 121
DB 189 NTSLSVRLKTIEMYRONVKKSDP-EVLFYQAYMLQATLTIESSNALVQDSKEGNVQS 247
QY 122 ---KSQKQKE-EAYLLFAKAADMNKLKAMEKADALLFGNFGVONITAAITOLYESLAKEG 177
DB 248 SLDLQFLKEAGSYL---KKLSKGYSDAQVLLADGYSAGFAFGKIEKAEFLVFOAAKHG 305
QY 178 SCAQNALGFLSYGIGMEYDQAKALIYYTFGSAGNMMSQMTLG-YYRLSGINVILQNC 236
DB 306 HIESAYRASHCLLEGLTTRDSKRSVNFELKFAASRNHPSAMYKGLYSFYGRMGLPTDVN 365
QY 237 VALSYKKVADYIADTFEKSQGVPEKVRILTERPENLSSNSELDDWDIYQYK-FLAERG 295
DB 366 TKLNGVKWLSRAARANELTAAPYELAKI-----YHEGFL----- 401
QY 296 DVQIQVSLGQLHLIGRGLDQDYKALHYFLKAAKAGSANAAMAFICKMYLEGNAAYPQNN 355
DB 402 DVVIP-----DEKY--AMELYIQNASLGHVPSATLLAQIYETGNDTVGQDT 445
QY 356 ATAFKYFSMAASKGNAIGLGLGLYFHG--KGVPLNYAEALKYFQKAAEKGPDAQFOL 413
DB 446 SLSVHYITQALKGDSVAMGLCAWYLLGAEPAFEDENEAFOWALRAANAGLPKQFTL 505
QY 414 GPMYSGSGIWKDKLAFYFYLASQSOGPLAI 446
DB 506 GFYFHEHGKCDRNMVEYAKWYKAGNEDKRAI 538

RESULT 5
YATA_SCHPO
ID YATA_SCHPO STANDARD; PRT; 932 AA.
AC Q09897;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 103.2 KDA PROTEIN C24B11.10C IN CHROMOSOME I.
GN SPAC24B11.10C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Odell C., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO YEAST SKT5, YEAST YER096W AND S.POMBE
CC SPCC417.03C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 267757; CAA91775.1; -
CC Hypothetical protein.
KW SEQUENCE 932 AA; 103183 MW; 86C1F6EFF94BB452 CRC64;

```



```

CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: CONTAINS 5 TPR REPEATS.
CC -!- SIMILARITY: BELONGS TO THE HCPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF008365; AAB63298.1; -
DR EMBL; AF053708; AAC24210.1; -
DR EMBL; AE000541; AAD07279.1; -
DR TIGR; HP0211; -
DR InterPro: IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.
KW Signal; Repeat; TPR repeat; Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 250 HCPA PROTEIN.
FT REPEAT 29 62 TPR 1.
FT REPEAT 67 98 TPR 2.
FT REPEAT 100 133 TPR 3.
FT REPEAT 133 169 TPR 4.
FT REPEAT 169 202 TPR 5.
FT VARIANT 4 4 N -> S (IN ATCC 49503).
FT VARIANT 9 9 L -> F (IN ATCC 49503).
FT VARIANT 11 11 G -> W (IN ATCC 49503).
FT VARIANT 17 17 T -> A (IN ATCC 49503).
FT VARIANT 91 91 G -> S (IN ATCC 49503).
FT VARIANT 194 194 K -> Q (IN ATCC 49503).
FT VARIANT 199 199 K -> N (IN ATCC 49503).
FT VARIANT 235 235 S -> G (IN ATCC 49503).
SQ SEQUENCE 250 AA; 27365 MW; A375FF12EB795355 CRC64;

Query Match 4.7%; Score 168.5; DB 1; Length 250;
Best Local Similarity 23.3%; Pred. No. 0.00032;
Matches 60; Conservative 34; Mismatches 102; Indels 61; Gaps 6;

QY 340 IGKMYLEGNAAPV-----QNNATAFYFESMAASKGNAIGLGLYFHG 384
Db 15 LGTILRLGLMAEPDAKELVNLGIESAKKQDFQAQAKELKGF 74
QY 385 KGVLNVAEALYFQAAEKGPDAQFOLGFMYSGGIWKYKLAFFYLASQSQPL 444
Db 75 KGVGKDLKKAIOFYTKGCELNDGCGNLLGNLYNGOGYSKDAKASQYYSKACDLNHA 134
QY 445 AIYLLAKMYATGTVVRSCTAVELYKGVCGLHWAKEKFLTAYFAYKD--GDIDSSLVQY 502
Db 135 GCMVLGSLHHYGVGTGPKDLRKALDLYEKACDL-----KDSPGCINAGYI-- 178
QY 503 ALLAEYGEVAQS--NSAFILSKKANILEKEMYPMALLLNRAAIOGNFAFVKIGDYH 561
Db 179 -----YSTKFKKEAIVRYSKACELKDGRCY-----NLGVMQ 211
QY 562 YYGYGTKKDYQTAATHY 578
Db 212 YNAQGTAKDEKQAVENF 228

RESULT 8
HCPA_HELPJ STANDARD; PRT; 250 AA.
AC Q9ZMM1;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HCPA PROTEIN PRECURSOR (CYSTEINE-RICH 28 KDA PROTEIN).
GN HCPA OR JHP0197.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

```

```

OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99120557; PubMed-9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 5 TPR REPEATS.
CC -!- SIMILARITY: BELONGS TO THE HCPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001458; AAD05781.1; -
DR InterPro: IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.
KW Signal; Repeat; TPR repeat; Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 250 HCPA PROTEIN.
FT REPEAT 29 62 TPR 1.
FT REPEAT 67 98 TPR 2.
FT REPEAT 100 133 TPR 3.
FT REPEAT 133 169 TPR 4.
FT REPEAT 169 202 TPR 5.
SQ SEQUENCE 250 AA; 27294 MW; 94EC37877E98C578 CRC64;

Query Match 4.5%; Score 161.5; DB 1; Length 250;
Best Local Similarity 23.0%; Pred. No. 0.00092;
Matches 62; Conservative 36; Mismatches 106; Indels 65; Gaps 7;

QY 332 GSNAMAF-----IGKMYLEGNAAPV-----QNNATAFYFESMAASKGNAI 372
Db 3 GSVKKTFLGVLGALGALRLGLMAEPDAKELVSLGIESVKKQDFQAQAKELKKEGF 62
QY 373 GLHGLGLLYPHGKGVPLNVAEALYFQAAEKGPDAQFOLGFMYSGGIWKYKLAFLK 432
Db 63 GCVFLGAFYEEGRGVGKDLKKAIOFYTKGCELNDGCGRLLGNLYNGOGYSKDAKASQ 122
QY 433 YFYLASQSQPLAIYLLAKMYATGTVVRSCTAVELYKGVCGLHWAKEKFLTAYFAYKD 492
Db 123 YYSKSELNHAECTVLGSLHHYGVGTGPKDLRKALDLYEKACDL-----KD 168
QY 493 --GDIDSSLVQYALLAEYGEVAQS--NSAFILSKKANILEKEMYPMALLLNRAAIOG 549
Db 169 SPCGINAGYM-----YGVAKNFEAIVRYSKACELKDGRCY----- 205
QY 550 NAFARVKIGDYHYGYGTTKDYQTAATHY 578
Db 206 -----NLGVMQYNAOGTAKDEKQAVENF 228

RESULT 9
YBET_ECOLI
ID YBET_ECOLI STANDARD; PRT; 184 AA.
AC P77296;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE HCPA PROTEIN PRECURSOR (CYSTEINE-RICH 28 KDA PROTEIN).
GN HCPA OR JHP0197.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

```



```
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -!- SIMILARITY: TO E.COLI YBEQ. SOME, TO YEAST SKT5 AND S.POMBE
CC SPAC24811.10C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000169; AAC73748.1; -
DR EMBL; U82598; AAB40848.1; -
DR EMBL; D90704; BAA35294.1; -
DR EMBL; D90705; BAA35299.1; -
DR Ecogene; Egl3650; yber.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 184 AA; 20915 MW; 7DC01030D545A0F9 CRC64;

Query Match
Best Local Similarity 4.4%; Score 158.5; DB 1; Length 184;
Matches 43; Conservative 34; Mismatches 55; Indels 23; Gaps 5;

QY 344 YLEGNAAVPPQNATAFKYFSAASKGNALGLHGLGLYFHGK-GVPLNVAEALKVFOKAA 402
DB 40 YLSSPGSLPFTTAAT-----ELSAIHGHTSQYRLGEFFLHSGDGKPLDYTORAYWEQSA 95
QY 403 EKGWPDAPQGLGFMYSSGIWKYKFLAFKPYFLASQSGQPLAIYLLAKMYATGTGVVRS 462
DB 96 EOENPRAQSKLGWIYKLGIVKPDTRKAILWYKEAAEQGYAHAQYTLGLIYRNGSGI--- 152
QY 463 CHTAVELKGVCELGHW-AEKF--LTAFAYKGDG 494
DB 153 -----NVNHYESQKWLKLTAKOHYKNAE 175

RESULT 10
YE28_METJA STANDARD; PRT; 567 AA.
ID YE28_METJA
```

```
Q58823;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1428.
GN MJ1428.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: CONTAINS 13 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67584; AAB99438.1; -
DR TIGR; MJ1428; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 9.
DR SMART; SM00028; TPR; 7.
KW Hypothetical protein; Repeat; TPR repeat; Complete proteome.
FT REPEAT 14 47
FT REPEAT 48 81
FT REPEAT 83 115
FT REPEAT 115 148
FT REPEAT 150 183
FT REPEAT 199 234
FT REPEAT 236 268
FT REPEAT 268 301
FT REPEAT 303 335
FT REPEAT 344 379
FT REPEAT 379 412
FT REPEAT 414 446
FT REPEAT 505 538
SQ SEQUENCE 567 AA; 66239 MW; 7CB4AC7210B7927E CRC64;

Query Match
Best Local Similarity 4.0%; Score 141.5; DB 1; Length 567;
Matches 112; Conservative 83; Mismatches 195; Indels 157; Gaps 27;

QY 22 KAEHNKROKERNVTTQVSVNEIKQYLHLEQRTSSNVINKRENLEKKKQKIRKG 81
DB 32 KAVECYLKALEKKNTPIDWFLAVLYHLEKYDSALEAIN--EALKISPSNIYFAYLKG 89
QY 82 I---QNKDILKKNHLOKQAEKNTDE-----GDQLFKMGKIVLQSQSQKQKEAYLL 133
DB 90 LIHYRGEITILAYK-YLKASEKIKNEELFEILDGTSVKYG-----RYEALKY 137
QY 134 FAKAADMGNLKAMEKMAA-----LLFGNFGVONITAAITOLYESLAKEG---SCK-----A 181
DB 138 YLKYKMANSKNLNLFKAGKIYLLFG-----DIDKADAFNEILOQNPNSHECKKIVECM 192
```

```
QY 182 QNALGFLSSYGIGMEVDQAKALIYYTFSGAGNMMSOMILGYRYLSGINVL-----QN 234
Db 193 EVNVNAINSY-----EDLNNGI-----TMKNKDYIGALKIFNKVLQIDEN 233
QY 235 CEVALSYKKVADYIDTFKSGVPEVKVRLTERPENLSSNSEIILDWDIYQYKFLAER 294
Db 234 SDISY-YKYSV---IAEIFEYK-----KALEYIDKSIFNRSLY-----YAKK 274
QY 295 GDVQIQVLSQGLHIGRGLDQDYKALHYFLKAAGKAGSANAMAFIGKMYLEGNAAPQON 354
Db 275 GDI-----LYKLG-----DEGAIEAYNAIKLNSONPAYFGLAILYRKGELEK 320
QY 355 NATAF-KYFSMAASKGNAIGLGLLYPHGK-----GVPLNAYAEALKYFO-----KAAEK 404
Db 321 SSNFQKVLTELEELSEEDISALNLYSLIGKAETTGIPKPYHEAMKYVDNLINLENSR 380
QY 405 GWPDQFQFGFMYSSGSIWKDYKLAFFYLASQSQPLAIYYLAKMYATGTGVVRSR 464
Db 381 WW-----YVRYGIYK-----LGNYKDAYESFMNA-----LRVNPKDISTLSKLA 420
QY 465 TAVELYKGVCELGHWAELKPLTAYFAFKDGDIDSSLYQYALLAEMGYEVAQSNFAFILESK 524
Db 421 IVLE-----KSGKIDEAITTYKILKI-----VNSLQFTCEID 453
QY 525 KANILEK 531
Db 454 --NILEK 458

RESULT 11
YES6_YEAST
ID YES6_YEAST STANDARD; PRT; 512 AA.
AC P39000;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 56.6 KDA PROTEIN IN RAD51-UBP9 INTERGENIC REGION.
GN YER096W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 249-454 FROM N.A.
RX MEDLINE=92257567; PubMed=1581961;
RA Shinohara A., Ogawa H., Ogawa T.;
RT "Rad51 protein involved in repair and recombination in S. cerevisiae
RT is a RecA-like protein.";
RL Cell 69:457-470(1992).
CC -!- SIMILARITY: SOME, TO YEAST SKY5 AND TO S.POMBE SPAC24B11.10C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18839; AAB64651.1;
CC DR EMBL; D10023; BAA20966.1; ALT_SEQ.
CC SGD; S0000898; SHC1.
```

```
KW Hypothetical protein.
FT CONFLICT 311 311 K -> R (IN REF. 2).
SQ SEQUENCE 512 AA; 56578 MW; 654D742482616658 CRC64;

Query Match 3.8%; Score 137; DB 1; Length 512;
Best Local Similarity 21.6%; Pred. No. 0.096;
Matches 109; Conservative 68; Mismatches 183; Indels 144; Gaps 25;

QY 43 ETQY-----LSHL-----EQTSSNV-----INKRENLEKKKNQIRIKGIO 83
Db 43 EVRQYSSVSSSHISRSQDVFTNDKERISSIGEDAMIDIDASPSLIEK-----YNSFP 94
QY 84 NKDILRNKHQKQAEKNFTDEGDLFKMGIKVLQSKSKQKE-----EAYLLEAK 136
Db 95 TRKILPE-----QDEFNVEDDASSLK-----EKSQSCSEIEIASEISEILNGT 141
QY 137 AADMGNLKAMERADALLFCNFG-----VQNTAAIQLYESLAKEGSKAQNALGFLSSYG 192
Db 142 SAD-GNSE-----FHDFAEPSPSQNESVALSFSQS-----NDLDFLNN-- 178
QY 193 IGMEYDQAKALIYYTFSGAGNMMSOMILGYRYLSGINVLONCEVALSYKKVADYIADT 252
Db 179 -----PSGSGSSNDIN-----RSTSSISLPHVSLDFNYSNLC--LTNE 216
QY 253 FEKSGVPEVKVRLTERPENLSSNSEILDW-DIYQYKFLAERGDIQVLSLQGLHLIGR 311
Db 217 VTASESHNVAKPHLGK-----NKKSLLPWKTIEMYGVEVVKTDIYNSFOYAYIL--R 270
QY 312 KGLDQDYKAL-----HYFLKAAK-----AGSANAMAFIGKMYLE 346
Db 271 VGLDTEKLHELKLEDESNFSVDSLKEYLVNDKAVILKLSAVGYPDAQYLLGDAYSS 330
QY 347 GNAAVPQNNATAPKYPESMAASKGNALGLGLGLLYPHGKGVPLNAYAEALKYFQKAEKGW 406
Db 331 GVFGKTKNR-RAPLLFSAAKRMHIESVYRTAICYEGGLGVTNRNAPKAVNFLTFAATKHH 389
QY 407 PQAQFQFGFMYSG-SGIWKD-----YKLAFFKYFLASQ--SGOPLAIYYLAKMYATG 456
Db 390 PAAMYKLGIVSYHGLMGLPDDILTKMDGYRWLRRAATSMASFFVCGAP---FELANIYMG 446
QY 457 TG--VVRSCRTAVELYKGVCELGH 478
Db 447 YKDLIISDPDYAMALYEKAAALGH 470

RESULT 12
ID RBPL_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBPL.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites";
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M88097; AAA29743.1; --  
HSSP; P36956; IAM9.  
Malaria; Receptor; Signal; Transmembrane.  
SIGNAL 1 17 POTENTIAL.  
CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.  
DOMAIN 18 2807 EXTRACELLULAR.  
TRANSMEM 2808 2826 POTENTIAL.  
DOMAIN 2827 2869 CYTOPLASMIC.  
SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).  
SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 3.8%; Score 136; DB 1; Length 2869;  
Best Local Similarity 20.0%; Pred. No. 1.2;  
Matches 155; Conservative 116; Mismatches 276; Indels 228; Gaps 36;

QY 17 TLTITKAEEHNRKQERNVTQVSNEIKOYLHSHILEQRTSSNV-----INKRENLEKK 71  
DB 1218 TLKSNAGKGENHDLLELNKTKGQMRDIYEKLKTAELKEGTVNELKDANEKANKVEPE 1277  
QY 72 NKORXI-----RI---KGTONKDIILKRN--KNHLOKQAEKFTDREGDOLFKNGI-KVLQQ 120  
DB 1278 PERNLIGHVLERITVEKDQKAGVVEEMSLTKIKELKIQEISDDSONELVTSITKHLEN 1337  
QY 121 SKS-----QKQKEAYLLFKAADMGNLKAMEKDALLFGNFGVONTAATAIQLYESLAK 175  
DB 1338 AKGYEDVIKRNEDSIQLREKAKSLTLDKMKLQVQV-----NMNLQSAIQ-----GN 1386  
QY 176 EGSCAQNALGFLSYGIGMEYDQAKALIYTFGAGNWSQMILGYRYLSGINVLQNC 235  
DB 1387 AGISKELNELKGVIELLSTNY--SSILEYVKKNSESVRFSQLANG-EFTKAEAGEEKNA 1443  
QY 236 EVALSVYKVV-----ADY--IADTFEKGSEVPYKVLTERPENLSNSILDWDIYO 286  
DB 1444 SARLAEAEKLKEQIVKDLIDYSDIDDKVKKIEGKIEILKME-----SALTWESE 1495  
QY 287 YYK-----FLAERGD-VQIQVSLGOLHLIGRKLGDODYY-KALHYF- 325  
DB 1496 KEQWCMSSHENAKEGKKIEYKLNKGDKGKANITDSOMEVGV-----NYVSKAEHAFH 1549  
QY 326 -----LKAAGKAGSANAMAFIGKM-----VLEGNAAVPQNNATAFKVFSAASKGNAI 372  
DB 1550 TVEAQVDTKAFCEISIVAYVTVMNDLNFESLMKEVKVKCEKKNDEAEKYSK----- 1601  
QY 373 GLRGLGLLYFHGK--GVPLNTYAEALKYFQKAAERGWPAQF-----QL 413  
DB 1602 -----LKPYDGRIKARVSENERKISELKEKAKVEKKESSQLNDVSTKSLQIDNCROOL 1655  
QY 414 GPMYSGGIWKDYKLFKIFYLASOSGOLAIYYLAKMYATGTGVVRSCTRAVELYKGV 473  
DB 1656 DSVL---SNIGRVKQNALQYFDSADKS-----MKSVLP 1686  
QY 474 CELGHWAKEFLTAYFAYKDG-DIDSSLYQYALLAEMGYEVAQNSAFILSK-----KANI 528  
DB 1687 SELG---AEKSLDKVKAAKESYEKNLETVO-----NEMSRINVEEGSLTDIDKKITDIENDL 1740  
QY 529 LEKEMYPMALLLNRAATOQNA-----FARVKGIDYHYYG 564  
DB 1741 LAMKQYEEGLL---QKIKENADRKRKNFELVGSFEINALLDPSISIFIKLALKEYDWTG 1796  
QY 565 -----YGTK-----KDYOPAAATHYSTAAN---KYHNAQAMFNLAYWYHGLGITKDI 608  
DB 1797 DLKNYGVKMKNEIGHEFTKSYNLIETHLSNATDYSVTFEKAQSLRELAKEEE----- 1848  
QY 609 HLARLYDMAAQTSPDAHPVLFAYWKLKLETHLLRDILFFNQFTTWNWKLKLDNT 663  
DB 1849 HLRRR-----EETAEIFLLNDIKKVBESLKLKEMM---KKVSAEYEGMKRDKHT 1892

## RESULT 13

EXOR\_RHLV STANDARD; PRT; 267 AA.  
AC Q52822;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE EXOPOLYSACCHARIDE PRODUCTION NEGATIVE REGULATOR PRECURSOR.  
GN EXOR.  
OS Rhizobium leguminosarum (biovar viciae).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=387;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WSM710 / WR6-35;  
RX MEDLINE=97346048; PubMed=9202471;  
RA Reeve W.G., Dilworth M.J., Tiwari R.P., Glenn A.R.;  
RT "Regulation of exopolysaccharide production in Rhizobium leguminosarum  
biovar viciae WSM710 involves exoR";  
RL Microbiology 143:1951-1958(1997)  
CC -!- FUNCTION: NEGATIVELY MODULATES EXOPOLYSACCHARIDE (EPS)  
CC BIOSYNTHESIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L39937; AAB63941.1; -- Signal.  
KW Exopolysaccharide synthesis; POTENTIAL.  
FT SIGNAL 1 23  
FT CHAIN 24 267 EXOPOLYSACCHARIDE PRODUCTION NEGATIVE  
FT REGULATOR.  
SQ SEQUENCE 267 AA; 28922 MW; E45E2E0538E1F573 CRC64;

Query Match. 3.7%; Score 133; DB 1; Length 267;  
Best Local Similarity 32.5%; Pred. No. 0.073;  
Matches 49; Conservative 20; Mismatches 62; Indels 20; Gaps 6;

QY 311 RKGLODDYYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVPQNNATAFKVFSAASKGN 370  
DB 55 KNGKEEAEVAYRAAEKGTGSRWALA---NNYADGD-GVTQDDDFEAFKIYSEIAQQGV 110  
QY 371 AIGLHGLGLL-----YFHG-KGVF-----LNYAEALKYFQKAAERGWPAQFQLGF 415  
DB 111 EPGSEDTGFFVNALLSLANYKKGAGSPVRIDLQAROL-YFOVASTGVPFAQFOLAQ 169  
QY 416 MYTSGGIWKDYKLFKIFYLASQSQGLPAI 446  
DB 170 NMLAGEGNGNASPQAKKWLNLQAKSGHPGAM 200

## RESULT 14

YG25\_HAEIN STANDARD; PRT; 165 AA.  
AC P44277;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN H11625.  
GN H11625.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2002, 15:03:36 ; Search time 19.39 Seconds  
(without alignments)  
2706.770 Million cell updates/sec

Title: US-09-714-882-2  
Perfect score: 3570  
Sequence: 1 MKPLSLLEILILGVTKT.....LFVIGLIVPGLILLRNHHG 689  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1159	32.5	685	2 T22223	sel-1 protein - Ca
2	723.5	20.3	604	2 E96762	hypothetical prote
3	455	12.7	700	2 T38028	hypothetical prote
4	391	11.0	807	2 T18454	hypothetical prote
5	373	10.4	833	2 S48558	probable membrane
6	314	8.8	327	2 B47999	hypothetical prote
7	275.5	7.7	713	2 T40051	hypothetical prote
8	267.5	7.5	256	2 G71856	hypothetical prote
9	258.5	7.2	256	2 E64659	conserved hypot
10	255.5	7.2	268	2 T46587	hypothetical prote
11	248	6.9	355	2 G71958	hypothetical prote
12	247	6.9	477	2 H81845	hypothetical prote
13	244	6.8	355	2 C64549	conserved hypot
14	240	6.7	468	2 F82012	hypothetical prote
15	228	6.4	195	2 F85564	hypothetical prote
16	228	6.4	467	2 B81095	conserved hypot
17	226.5	6.3	290	2 C71859	hypothetical prote
18	222.5	6.2	290	1 B84657	conserved hypot
19	219	6.1	696	2 S39827	SKT5 protein - yea
20	214	6.0	306	1 H64539	conserved hypot
21	214	6.0	352	2 G71914	hypothetical prote
22	212	5.9	198	2 F81200	conserved hypot
23	210.5	5.9	305	2 G71968	hypothetical prote
24	200.5	5.6	623	2 S67762	hypothetical prote
25	195.5	5.5	364	2 A96730	hypothetical prote
26	187	5.2	328	2 F83599	hypothetical prote
27	183.5	5.1	251	2 F71945	hypothetical prote
28	183.5	5.1	512	2 T41340	hypothetical prote
29	180	5.0	932	2 S62555	protoplast regene

30	176.5	4.9	225	2 D64598	hypothetical prote
31	170.5	4.8	321	2 T47138	hypothetical prote
32	169	4.7	271	2 F83081	hypothetical prote
33	168.5	4.7	250	2 C64546	conserved hypot
34	168	4.7	2295	2 B71621	probable membrane
35	162	4.5	256	2 E71807	hypothetical prote
36	161.5	4.5	250	2 B71961	hypothetical prote
37	158.5	4.4	184	2 E84799	yber protein - Esc
38	156.5	4.4	184	2 B85565	hypothetical prote
39	153	4.3	633	2 T39352	hypothetical prote
40	153	4.3	1115	2 B84476	probable TPR repea
41	147	4.1	456	2 T40386	hypothetical prote
42	145.5	4.1	289	2 A83011	hypothetical prote
43	141.5	4.0	385	2 A82388	conserved hypot
44	141.5	4.0	567	2 C64478	hypothetical prote
45	140.5	3.9	138	2 H64561	invalid gene - Hel

ALIGNMENTS

RESULT 1

T22223  
sel-1 protein - Caenorhabditis elegans  
N:Alternate names: regulator of lin-12  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T22223; S68303; S68304  
R:White, S.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19532  
A:Accession: T22223  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-685 <WIL>  
A:Cross-references: EMBL:Z78063; PIDN:CAB01505.1; GSPDB:GN00023; CESP:F45D3.5  
A:Experimental source: clone F45D3  
R:Grant, B.; Greenwald, I.  
Genetics 143, 237-247, 1996  
A:Title: The Caenorhabditis elegans sel-1 gene, a negative regulator of lin-12 and gl  
A:Reference number: S68303; MUID:96304591  
A:Accession: S68303  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-685 <GRA>  
A:Cross-references: GB:U50828; NID:g1255198; PIDN:AAC47112.1; PID:g1255199  
A:Accession: S68304  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-685 <GR2>  
A:Cross-references: EMBL:U50829; NID:g1255200; PID:g1255201  
C:Genetics:  
A:Gene: sel-1  
A:Map position: 5  
A:Introns: 41/1; 90/2; 210/3; 411/1; 548/1

Query Match 32.5%; Score 1159; DB 2; Length 685;  
Best Local Similarity 38.7%; Pred. No. 1.2e-64;  
Matches 253; Conservative 137; Mismatches 214; Indels 50; Gaps 11;

QY	53	EORTSSNVINKRENLEKKNKIRIKIQNKDILK-RNKNHLOKQAKENFTDEGDQLF	111
Db	57	EQVTSQDENK-----SNREIPKIVISEEYLAEKVEQPTSPAEAEF	97
QY	112	KMGIRVLOQSK--SQKQKEAYLLFAKADMGKLNKAMEKMDALLFNGFVGQVNTAIQL	169
Db	98	QGMAYIERGKGHREGRVAAHRVFERAAQHQEARKAVAFSQMGFDSRWSIQEAKTV	157
QY	170	YESLAKEGSCKAQNALGFTLSSYGIGME-YDQAKALIYTFGSAGGNMWSOMILGYRLSG	228
Db	158	FEDLEKNGSPDQALALGFMHMGAGIGVEKSNQAKALVYWFSLGNGPLAQMAMGFRIYSHG	217





```
QY 387 VPLNYAALKYFQKAAKGWPDQAQFQGMYYSG-----SGIWKD----- 426
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 353 PLFDDIDKAYWFKRGATKNDNSYYGLGYMAYHGLTNGVDREKGMRLINLAVMNNPHA 412
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 -----YKLAIFYFYLASOGOPALAIYYLAKMYATGTGVVRSCTRAVELYKGV 473
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 LMFGLGIRLEEARIEAYHFLRAATQKSVISYKYLADCYNGTGTSRSMISASLYYK-- 470
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 474 CELGHWAERFLTAYPAY-----KDGIDSSLYOYALLAEMGYEVAQSNFAFIL 521
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 471 -----KFVEAIRASATSMALAEIDEYGFHNSFYVLYAAQMGYALAEINAYILM 522
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 522 ESKKANI-----LEKEKMY-PMALLLNRAAIQGNFAFARVKIGDIHYHYGYGTRK 569
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 DENKFLINSVFYFNYSQSEAAHDKFAEFYSRAAAGQDIDAIFKLGDIYYHYGIGTPK 582
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 570 DYQTAATHYSIAANKYHNAQMFNLAYMEHGLGITKDTHLARLYDMAAQSPDAHIPV 629
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 583 DYSKAYTCYKIAEQSSIGMGLWNMAYMEHYGIGRQDDIYIARRLLD-ELSSNQNSYFPL 641
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 630 LPVAMKLETHLLRDIL 646
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 642 KVAIFWINIHQLYIKLL 658
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
T18454
hypothetical protein C0550w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18454
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18937
A:Accession: T18454
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-807 <LAW>
A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332547; PIDN:CAA15596.1
C:Genetics:
A:Map position: 3
A:Introns: 512/3; 638/1; 750/2
A:Note: C0550w

Query Match 11.0%; Score 391; DB 2; Length 807;
Best Local Similarity 22.1%; Pred. No. 1.1e-16;
Matches 152; Conservative 106; Mismatches 237; Indels 194; Gaps 20;

QY 24 EEHNKROKERNVTVQSVNEIK--QYLSHILEQRTSSNVINKRENLEKKKNQKIRIKG 81
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 EEEKKKKLYLKNVDRANDLDSFSYIHNNEDSSNDNNGNKYNNRKNQNDNNKGD 151
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 IQNKDIL--KRN-----KNNHLOKQAEKNF 103
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 YNNKGYNNKRNYSKONDKDEYNKYTYDDNNDYNNNEENYNDYNEINKEKEKEK 211
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 TDEGDOLFNGIKVLOQSKOK--EAYLLFAKADMGNLK---AMEKMDALLFG-- 156
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 EDRKQWFKLAIEBKNGSKKKKNINKCIEIFOKLITDKNDKITSYSELGYIFEGYK 271
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 NFGV---QNTAAIQYLSLAKEGSCQAQNALGLFLSSYGI----- 193
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 NYFFSKRNVLNLHLVLRQAAMKNPAALHFLSFIFYFDHPKTEENKRNKKYIQNVQN 331
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 194 -----GNEYDQ-----AKALIYYTSGAGNMMSQMLGYRSLGNLQNCVAL 239
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 332 VNNNNINYQNLOQNFNFKKSIEFEMIAASLNYPISLITLTKYKFLYGNKQNCYKAK 391
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 SYKKVA-----DYLTADTEKSEGVPVEKVRTERPENLSSNSEILDWDIYYKFLAE 293
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 392 KLYKNAENVNNSDYI-----NIPLSELDLL-NGENLNMEIN----- 428
```

```
QY 294 RGDVQIQVSLGSLHLIGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAPVQ 353
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 -----NMKNNEEIELEFNEQIKGGVMAVMDLGRKYE-----EK 465
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 354 NNATAPKYSMAASKNAIHLGLGLLYPHGKGVPLNYAEALKYFOKAAEKGWPDQAQFL 413
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 466 NFKQAFKYINEASKNNLALRELGIYLYGYGVQDINKSIENFSAKAAEAGDVESKCYL 525
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 414 GMVYSGSGIWKDYKLAIFY-----FYLA----- 437
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 GTIYYPIDG-YKNLESLKYLIEAASHDYGAEFFFLAEIILDISMRKQVYISDYVYEVFK 584
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 438 -----SOSGOLAIYYLAKMYATGTGVVRSCTRAVELYKVCGLHWAERFLTAIFYKD 492
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 585 LYEHSDADLVGVQAYPREAQLYEIGKGVKQSCNLATLSYKFAESTLWINNIRQGM DYILE 644
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 493 GDIDSSLYOYALLAEMGYEVAQSNFAFILESKKA-NILEKEKMYPMALLLNRAAATQNA 551
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 645 KDYLKAFYTYALASYEGYEIAQNVLNLYIYRTNKLNNVIHPRKI-----MLVNLNLYKOGNY 700
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 552 FARVKIGDYH-----YYGYGTTK 569
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 701 KALYEMGEIYKEONKEELSVSYYKLGKK 729
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
S48558
probable membrane protein YLR207w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L8167.5
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 06-Feb-1998
C:Accession: S48558
R:Pauley, A.
submitted to the EMBL Data Library, September 1994
A:Description: The sequence of S. cerevisiae cosmid 8167.
A:Reference number: S48545
A:Accession: S48558
A:Molecule type: DNA
A:Residues: 1-833 <PAU>
A:Cross-references: EMBL:U14913; NID:g544497; PID:g544502; MIPS:YLR207w
C:Genetics:
A:Gene: SGD:HRD3
A:Cross-references: SGD:S0004197; MIPS:YLR207w
A:Map position: 12R
C:Keywords: transmembrane protein
F:2-18/Domain: transmembrane #status predicted <TM1>
F:768-784/Domain: transmembrane #status predicted <TM2>

Query Match 10.4%; Score 373; DB 2; Length 833;
Best Local Similarity 23.9%; Pred. No. 1.5e-15;
Matches 165; Conservative 106; Mismatches 250; Indels 168; Gaps 28;

QY 97 KQAEKNF-----TDEGDOLFNGIKVLOQSKSQKQKEEAYLL-----FAKAA 138
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 RNEENYQSIQWNEITDSQRHIYEL---LVQSEQNNSSEATYTLQIHLWSQYNFPHNM 124
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 139 DMGNLAKEMKADALLFGNFGVNITAAIQYLSLAKEGSCQAQNALGFLSSYIGMEYD 198
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 TLAH-KYLEKFNDLTHTNH-----SAIFDLAVWATGCGASGNDQTVIPQ-----D 170
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 199 QAKALIYYTSGAGNMMSQMLGYRSLGNLQNCVALSYKKYKVDYIADTFEKE- 257
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 SAKALLYQRAAQLGNLKAQVLAIFYSGFNVPNPFKSLVLYRDAEQLRKSYSRDEW 230
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 -----GVPVEKVR-LTERPENLSSNSEILDWDIYYK 289
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 DIVFPYTESYNNRISDFESGLGKGLNSVPSVTVRKTRTPDIGS-----P 276
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 FLAERGDVQIQVS---LGQLHLIGRKGL-----DQD-----YKALHYFLKAAKA 331
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 FLAQVNGVQMTLQIEPMGRFAFNQDNGNNGDEDEDEDASERRIRIYYAALNDY-KGTYS 335
```









---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2002, 15:04:26 ; Search time 31.14 seconds  
(without alignments)  
3236.404 Million cell updates/sec

Title: US-09-714-882-2  
Perfect score: 3570  
Sequence: 1 MKPLSLLEILLIGVTIKT.....LFVIGLIVPLGILLRRHHG 689

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2650.5	74.2	505	4 Q9UGD3	Q9ugd3 homo sapien
2	1640	45.9	794	11 Q9ESM7	Q9esm7 mesocricetu
3	1264	35.4	819	5 Q9V415	Q9v415 drosophila
4	1159	32.5	685	5 Q20423	Q20423 caenorhabdi
5	1156	32.4	685	5 Q17378	Q17378 caenorhabdi
6	817	22.9	678	10 Q9LM25	Q9lm25 arabidopsis
7	723.5	20.3	604	10 Q9C6B6	Q9c6b6 arabidopsis
8	455	12.7	680	3 Q13875	Q13875 schizosacch
9	435	12.2	1029	4 Q9A847	Q9a847 homo sapien
10	435	12.2	1201	2 Q9XDH8	Q9xdh8 legioneella
11	421.5	11.8	392	10 Q9FX49	Q9fx49 arabidopsis
12	391	11.0	807	5 Q77341	Q77341 plasmodium
13	373	10.4	833	3 Q05787	Q05787 saccharomyc
14	349	9.8	71	4 Q9NPY6	Q9np6 homo sapien
15	294.5	8.2	469	2 Q9RN76	Q9rn76 coxiella bu
16	275.5	7.7	713	3 Q9USV0	Q9usv0 schizosacch
17	271.5	7.6	256	2 P94848	P94848 helicobacte
18	267.5	7.5	256	2 Q9ZK96	Q9zk96 helicobacte
19	258.5	7.2	256	2 Q25742	Q25742 helicobacte

20	255.5	7.2	268	2	Q9RBX2	Q9rbx2 pseudomonas
21	248	6.9	355	2	Q9ZMJ9	Q9zmj9 helicobacte
22	247	6.9	477	2	Q9JU20	Q9ju20 neisseria m
23	244	6.8	355	2	Q25021	Q25021 helicobacte
24	240	6.7	468	2	Q9JWX2	Q9jwx2 neisseria m
25	232	6.5	172	2	Q9CKK3	Q9ckk3 pasteurella
26	228	6.4	467	2	Q9JZ25	Q9jz25 neisseria m
27	226.5	6.3	290	2	Q9ZKB5	Q9zkb5 helicobacte
28	222.5	6.0	320	2	Q25728	Q25728 helicobacte
29	215.5	6.0	330	2	Q9A6P5	Q9a6p5 caulobacter
30	214	6.0	306	2	Q24968	Q24968 helicobacte
31	214	6.0	352	2	Q9ZLX8	Q9zlx8 helicobacte
32	212	5.9	198	2	Q9K0Y5	Q9k0y5 neisseria m
33	210.5	5.9	305	2	Q9ZMS0	Q9zms0 helicobacte
34	200.5	5.6	623	3	Q07622	Q07622 saccharomyc
35	199.5	5.6	242	2	Q9CNO9	Q9cng9 pasteurella
36	197	5.5	201	10	Q40097	Q40097 ipomoea tri
37	195.5	5.5	364	10	Q9S730	Q9s730 arabidopsi
38	195.5	5.5	392	2	Q66397	Q66397 acinetobact
39	194	5.4	95	4	Q12916	Q12916 homo sapien
40	189.5	5.3	687	3	P87065	P87065 candida alb
41	187	5.2	328	2	Q9I6D4	Q9i6d4 pseudomonas
42	183.5	5.1	251	2	Q9ZMA4	Q9zma4 helicobacte
43	183.5	5.1	512	3	Q94486	Q94486 schizosacch
44	183	5.1	510	11	Q9DCV6	Q9dcv6 mus musculu
45	176.5	4.9	225	2	Q25345	Q25345 helicobacte

## ALIGNMENTS

RESULT 1

Q9UGD3	PRELIMINARY;	PRT;	505 AA.
ID	Q9UGD3		
AC	Q9UGD3;		
DT	01-MAY-2000 (TREMBlrel. 13, Created)		
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)		
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)		
DE	DJ842G6.2 (NOVEL PROTEIN IMILAR TO SEL1L (SEL-1 (SUPPRESSOR OF LIN-12, C-ELEGANS)-LIKE)) (FRAGMENT).		
DE	C-ELEGANS)-LIKE)) (FRAGMENT).		
GN	DJ842G6.2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Barlow K.;		
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL109657; CAB65792.2; -		
FT	NON_TER		
SQ	SEQUENCE 505 AA; 57003 MW; 536D4361FA826E35 CRC64;		
Query Match	74.2%;	Score	2650.5; DB 4; Length 505;
Best Local Similarity	99.8%;	Pred. No.	1.7e-163;
Mismatches	505;	Conservative	0; Mismatches 0; Indels 1; Gaps 1;
QY	184	ALGFLSSYGIGMEYDQAKALIYYTFGSAGNNMMSQMLGYRLSGINVLQNCNEVALSYK	243
Db	1	ALGFLSSYGIGMEYDQAKALIYYTFGSAGNNMMSQMLGYRLSGINVLQNCNEVALSYK	60
QY	244	KVADYIADTFEKGVPVKEKRLTERPENLSSNSEILDWDIYQYKFLAERGQVQVSL	303
Db	61	KVADYIADTFEKGVPVKEKRLTERPENLSSNSEILDWDIYQYKFLAERGQVQVSL	120
QY	304	GQLHLIGRGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVPQNNATAFKYES	363
Db	121	GQLHLIGRGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVPQNNATAFKYES	180
QY	364	MAASKGNATGLHGLGLLYFHGKGVPLNYAALKYFKAAEKGWPDQAFQGLGFMYSGSGI	423
Db	181	MAASKGNATGLHGLGLLYFHGKGVPLNYAALKYFKAAEKGWPDQAFQGLGFMYSGSGI	240







QY 522 ESKANIL---EKKMYPMALLNNRAALQGNAPARVIGDYHYGYGKDYQTAATHY 578  
Db 518 DRGEATSLFSGPKDNMERAFLNQRSANQYAAARVLDGDIYYGLGTEVDHSLAFSNY 577  
QY 579 STAAKNYHNAQAMFNLAYMYEHLGITDKIDHLARLYDMAAQTSPDAHIPVLFAYMKLET 638  
Db 578 KMAVDXHGVAQAMFNLYGMYHVEGEGITRDLYLAKRFDQAIEHSODAYMPSKALAKLAF 637  
QY 639 THLLRDILFFNQFTTRWNWKL-----DNITGPHWD--LFVIGLIVPGLILLRN 686  
Db 638 VFYLEEL-----NKLPLISPMEKTGVRWDAILMTVSALVP-LFLFWRH 680

RESULT 5  
Q17378 PRELIMINARY; PRT; 685 AA.  
AC Q17378;  
DT 01-NOV-1996 (TremBLrel. 01, Created)  
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)  
DE SUPPRESSOR/ENHANCER OF LIN-12 (SEL-1).  
GN SEL-1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N2;  
RX MEDLINE=96304591; PubMed=8722778;  
RA Grant B., Greenwald I.;  
RT "The Caenorhabditis elegans sel-1 gene, a negative regulator of lin-12  
RT and gln-1, encodes a predicted extracellular protein.";  
RL Genetics 143:237-247(1996).  
DR EMBL; U50828; AAC47112.1; -  
SQ SEQUENCE 685 AA; 76222 MW; 3779A39937A4FC52 CRC64;

Query Match 32.4%; Score 1156; DB 5; Length 685;  
Best Local Similarity 38.7%; Pred. No. 1.2e-66;  
Matches 253; Conservative 136; Mismatches 215; Indels 50; Gaps 11;  
QY 53 EORTSSNVINKRENLEKKKQIRIKIQNKDILK-RNKNHLOKQAEKNFTDEGDQLF 111  
Db 57 EQVTSQDENK-----SNREIPKVISEYLAKEVQPPSPAEAEF 97  
QY 112 KMGIVLOQSK--SOKQEEAYLLFAKADMGNLKAMEKMDALLFGNFGVONTAAIQL 169  
Db 98 QRMAYIERGKGHGRVAAHRVFERAAAOHQEARKAVAFSOMFGDYSRWSIOEAKTV 157  
QY 170 YESLAKEGCKAQNALGFLSSVIGME-YDOAKALIYKTFGSAGGNMMSOMILGYRILSG 228  
Db 158 FEDLEKNSPDQAALGFMHMGAGIVKESNQAKALVITFSAAGNPLAQAMANGFRYSHG 217  
QY 229 INVQNCVALSYKRVADYIADTFEKGSEGVPEKVRITERPE---NLSSNSEILDWIIY 285  
Db 218 VGPQNCETALSYYQKAVTVVDNVKFTTGQIQLRLTDETDPTIHQPGSAPLESNLL 277  
QY 286 QYVFLAERGVDQIVQSGOLHIGKGLDQDYKALHYFLKAAGSAGNAMAFTGKMYL 345  
Db 278 EYKMLADKGDTSAGLGLQIYLAGRGLNQNFELAFRYLLAAEESGADALTYLGKMYL 337  
QY 346 EGNAAVPPONNATAFKYSMAASKNAIGLHGLGLYFHCKGVPLNAYEALKYFQKAAEKG 405  
Db 338 DGTPTPKDYQKSFELMKSDKSPSAQAVLGMKGVKKKYNKALKLTLTSDAKK 397  
QY 406 WPDQFQGFMYISG-----SGTWKYKIAFKYFLASQGPPLAIYILAKMYATGTGYVR 461  
Db 398 NADGOMYLAELHYKVPYKNGVHRDFKSKVLYQLASQNGHILAYNLQAQMHAGTGYVR 457  
QY 462 SCRTAVELYKGVCELGHNAEKFLTAIFYAYKDGIDSSLVQYALLAEMGYEVAQNSAFLL 521  
Db 462 SCRTAVELYKGVCELGHNAEKFLTAIFYAYKDGIDSSLVQYALLAEMGYEVAQNSAFLL 521

Db 458 SCSHAVDLFKSVAERKGRMEASAYKONRVDEAAAMKYLFAELGYEVAQTNLAYIL 517  
QY 522 ESKANIL---EKKMYPMALLNNRAALQGNAPARVIGDYHYGYGKDYQTAATHY 578  
Db 518 DRGEATSLFSGPKDNMERAFLNQRSANQYAAARVLDGDIYYGLGTEVDHSLAFSNY 577  
QY 579 STAAKNYHNAQAMFNLAYMYEHLGITDKIDHLARLYDMAAQTSPDAHIPVLFAYMKLET 638  
Db 578 KMAVDXHGVAQAMFNLYGMYHVEGEGITRDLYLAKRFDQAIEHSODAYMPSKALAKLAF 637  
QY 639 THLLRDILFFNQFTTRWNWKL-----DNITGPHWD--LFVIGLIVPGLILLRN 686  
Db 638 VFYLEEL-----NKLPLISPMEKTGVRWDAILMTVSALVP-LFLFWRH 680

RESULT 6  
Q9LM25 PRELIMINARY; PRT; 678 AA.  
AC Q9LM25;  
DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
DE T10022.22.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,  
RA Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
RA Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC T10022 from chromosome  
RT I.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC069551; AAF78381.1; -  
SQ SEQUENCE 678 AA; 75941 MW; 415235B57FC7430D CRC64;

Query Match 22.9%; Score 817; DB 10; Length 678;  
Best Local Similarity 36.6%; Pred. No. 9.6e-45;  
Matches 189; Conservative 99; Mismatches 185; Indels 44; Gaps 12;  
QY 137 AADMGNLKAMEKMDALLFGNFGVONTAAIQLYESLAKEGCKAQNALGFLSSVIGM- 195  
Db 99 AASEGNFLMEEAVDEI-----EASSAGDPHAQSTMGFV--YGIGMM 139  
QY 196 EYDOAKALIYKTFGSAGGNMMSOMILGYRILSGINVLQNC-EVALSYKKVADYIATF 253  
Db 140 REKSKSFLHNFHFAAGNMOSKMAFYLR-----ODMHDKAVQLYAELEAVNSF 194  
QY 254 EKSEGVV-VEKVRV---TERPENLSSNSEILDWIIYQYKFLAERGVDQIVQSGOLHIL 309  
Db 195 LISKDSPVVEPRIHSGTEENKALRKSGEDEDFQILEYQAKGNANAMWKIGLYFF 254  
QY 310 GRKGLDQDYKALHYFLKAAGSAGNAMAFTGKMYLEGNAAPVQNNATAFKYSMAASKG 369  
Db 255 GLRGLRDRHTKALHFLKAVDKGEPRSMELLGELIYARG-AGVERNTRYALEMLTAAKEG 313  
QY 370 NAIHLHGLGLYFHCKGV-PLNAYEALKYFQKAAEKGWPDQFQGFMYSSSGITWKDYK 428  
Db 314 LYSAPNGYLYKVGVDKKNYTKAREYFEKAVONEDPSGHYNGLVLYLKIGVNRDVR 373  
QY 429 LAFKYFYLASQSGQPLAIYILAKMYATGTGYVRSCRTAVELYKGVCELGHNAEKFLTAIF 488  
Db 374 QATKFFVNAAGQKAFYQLAKMFTGVGLKKNLEMATSFYKLVARCPWNSLSRWALE 433  
QY 489 AYKDGIDSSLVQYALLAEMGYEVAQNSAFILSKKANIL-----EKEKMPMAL 539

```

Db 434 AYLGKGVGRKALILYSMAEMGEVAQSNAAWILDKYGRSCMGVSGFCTDKER-HERAH 492
QY 540 LLWNRRAAIOGNFARVKIGDYHYGYGKDYQTAATHYSIAANKYHNAQAMFNLAYVE 599
Db 493 SLWWRASEQNEHAALLIGDAYIYGRGTERDFVRAEAY-MHAKSOSNAQAMFNLGYMHE 551
QY 600 HGLGITKDIHLARRLYDMAAQTSPDAHIPVLFAV 636
Db 552 HQGLPFDLHLAKRYIDYDESLSQDAARLPVTALASL 588

RESULT 7
Q9C6B6 PRELIMINARY; PRT; 604 AA.
AC Q9C6B6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 67.8 KDA PROTEIN.
GN F6D5.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltscher J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Aker C.M., Venter J.C., Davis R.W.
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
DR EMBL; AC079676; RAG51809.1; -.
KW Hypothetical protein.
SQ SEQUENCE 604 AA; 67761 MW; 973AFDAE38D7EBA6 CRC64;

Query Match 20.3%; Score 723.5; DB 10; Length 604;
Best Local Similarity 34.4%; Pred. No. 9e-39;
Matches 177; Conservative 90; Mismatches 160; Indels 87; Gaps 12;

QY 137 AADMGNLKAEMKADALLFGNFGVQNTAAIQLYESIAKEGSKAGNALGFLSSYGIGM- 195
Db 100 AASEGNTLME-----AVSEIDSSAGSGPHQAQSVMGFV--YIGWM 140

QY 196 -EYDQAKALIYTFSGAGNMKSOMILGYRYSGLNVLQNC-EVALSYKKVADYIADTF 253
Db 141 RETSRKSLIHFFHAAGNMQSKMALAFRYLR-----QNMVDKAVELAEIAETAVNSF 195

QY 254 EKSEGVV-EKVL-----TERPENSSNSELWDIYQYKFLAERGDVQIQVSLQHLI 309
Db 196 LISKDSPMAEPVRIHTGTEENKDALRSGEDEDFOILEYQAEKNSVAMHKGIFLYF 255

QY 310 GRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMWLEGNAVPONNATAFKYFSMAASKG 369
Db 256 GLRGLRDHAKALYWFSKAE----- 275
```

```

QY 370 NAIGHLGLLHYFHGKV-PLNYAEALKYFQKAAEKGWPDQAFQGLGFMYSGSGIWKDYK 428
Db 276 ---ENGLGYLVKGYGVGDKRNYTKAREYFEMAANNEDPSGHYNLGLVLYKGTGVKKDVR 331
QY 429 LAFKYFYLASOGQPLAIYYLAKYATGTGVVRSCTRAVELYKGYVCELGHNAKEFLTAYF 488
Db 332 HATKYFFVAANAGOPKAFYQLAKMFHTVGLTKNEMATTFYKLVRAERGPWSLSRWALE 391
QY 489 AYKGDGIDSSLVQYALLAEMGYEVAQSNASFTLESKANIL-----EKEKMPMAL 539
Db 392 AYLGKGVGRKALILYSMAEMGEVAQSNAAWILDKYGRSCMGVSGFCTDKER-HDRAH 450
QY 540 LLWNRRAAIOGNFARVKIGDYHYGYGKDYQTAATHYSIAANKYHNAQAMFNLAYVE 599
Db 451 SLWWRASEQNEHAALLIGDAYIYGRGTERDFVRAEAYMYAKSQ-SNAQAMFNLGYMHE 509
QY 600 HGLGITKDIHLARRLYDMAAQTSPDAHIPVLFAV 633
Db 510 HQGLPFDLHLAKRYIDYDALQSDTAAKLPVTAL 543

RESULT 8
Q13875 PRELIMINARY; PRT; 680 AA.
AC Q13875;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 78.7 KDA PROTEIN C1B3.10C IN CHROMOSOME I PRECURSOR.
GN SPAC1B3.10C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO C.ELEGANS SEL-1.
DR EMBL; Z98598; CAB11247.2; -.
KW Hypothetical protein; Signal.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 680 HYPOTHETICAL PROTEIN C1B3.10C.
FT DOMAIN 573 576 POLY-TYR.
SQ SEQUENCE 680 AA; 78703 MW; 522BD933D91FEC9 CRC64;

Query Match 12.7%; Score 455; DB 3; Length 680;
Best Local Similarity 26.3%; Pred. No. 2.5e-21;
Matches 162; Conservative 97; Mismatches 218; Indels 140; Gaps 21;

QY 126 QKEEAYLLFAKADMGNLKAMEKADALLFGNFG-VQNTAAIQLYESIAK-EGSKAQN 183
Db 86 QRNHATELLRSASHDNTDAMLYLANIEFFGLFEIPEIETSDSVKYDYMLQKANGSAFANN 145

QY 184 ALGFLSSYGIGMEY---DOAKALIYTFSGAGNMKSOMILGYRYSGLNVLQNC-EVALS 240
Db 146 MNGFYTSYTSFS-EYASNNPALARIHWELAAKQSGDADHQLAYHNLIALNMPQSDAEAYK 204

QY 241 YKKVADYIADTFEKSEGVVPEKVRLETERPENSSNSELID--WDIYQYKFLAERGDVQ 298
Db 205 HYKFIISDLH---FE-----EECGSNVTYKLCIWPIDQYENAGENG--- 242

QY 299 IQVSLGQLHLIGRGLDDQY-----YKALH-----AAKAGSANAMAFIGKMWLE 346
Db 243 -----MGVYGAASATYTSDAYQALHTRSOYLREMSNSTEDWDYELMPEVAKLRH 292

QY 347 GNAVPONNATA-----FKYFS-----MAASKGNAIGHG-----LGLLPHGKG 386
Db 293 GMYKPRNRYTSDVLFKRSQRYWPTYTSNSVLANTPQSIILAAQSCGYLGLLHFLDKG 352
```

```
QY 387 VPLNYAEALYFOKAAEKGWPDQFOLGFMYYSG-----SGIWKD----- 426
Db 353 PLEFDIDKAYWFRGATKNDNSYVGLGYMAYHGLTSNGVDREKGRMLNLAVNMENPHA 412
QY 427 -----YKLAIFYLASOGOPALAIYIYAKMYATGTGVVRSRTAVELYKGV 473
Db 413 LMFGLGIRLEAREYAYHLFLRAATQKSVISYKYLADCYNGTGTSRSMISASLYK-- 470
QY 474 CELGHWAKEKLTAVFAY-----KGDIDDSLVQYALLAEMGYEVAQSNFAL 521
Db 471 -----KEVEAIRASATSMATALEIDEYGFHNSFVYLYAAQMGVALAEINAYLM 522
QY 522 ESKKANI-----LEKERMY-PMALLLNWRAAIOGNAPARVKIGDYHYGYGTRK 569
Db 523 DENKELINSVERVFNVTQSEAAHDKFAVEFYSRAAQGDIDAIFKLGDYGYIGTRK 582
QY 570 DYOTAATHYSIAANKYHNAQAMFNLAHYEHGIGITKDIHLARRLYDMAAQTSPDAHIPV 629
Db 583 DYSKATCYKIAYEQSSIGMGLNMAHYHEYGIGRQDIYIARRLLD-ELSSNQNSYFPL 641
QY 630 LFVAMKLETHLLRDIL 646
Db 642 KVAIFWINHQLIKLL 658

RESULT 9
O94847 PRELIMINARY; PRT; 1029 AA.
AC O94847;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE K1AA0746 PROTEIN (FRAGMENT).
GN K1AA0746.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:277-286(1998).
DR EMBL; AB018289; BAA34466.1; -.
FT NON_TER 1
SQ SEQUENCE 1029 AA; 117551 MW; 1A38EE1162E79ACD CRC64;
```

Query Match 12.2%; Score 435; DB 4; Length 1029;  
Best Local Similarity 27.1%; Pred. No. 8.9e-20;  
Matches 160; Conservative 101; Mismatches 254; Indels 76; Gaps 20;

```
QY 100 EKNFTDEGDOLF---KMGIKVLQSKSQKQKEAYLLFAKADMGNLKAMEKMDALLF 155
Db 397 ERELKDKHPSLFQALLEMDLLTVPRNQNESVSIGKIFKA--VKRLSSID----- 446
QY 156 GNFQVONITAAIQLESRAKESCKAQNALGFLSSYGIGMEYDQAKALIYTFGSAGNM 215
Db 447 ---GLHQISSIVPFLDSCCGVHKASYVLAVFYETGLNVPDQGLMYSLVGGGSR 503
QY 216 MSQMILGYRLSGI-NVLQNCVALSYKKVADYIADTFEKSEG--VPVEKRLTLRPERN 272
Db 504 LSSMNLGYKHQYIDNYPDLMDWLSYAYSNIATKTPDQHTLQGGQAYVETIRL----- 557
QY 273 LSSNSIL-----DWDIYQYKFLAERGDVQIQVSLGQLHLIGKLDQDYKALHYFL 326
Db 558 --KDDEILKVQTKEDGVFPMWLKHEATRGNAQAQORLAQMLFWGQGVAKNPEATEWYA 615
```

```
QY 327 KAAKAGSANAMAFICKMYLEGNAAPQNNATAFKYFSMAASKGNAIGLHGLGLLYFHGK 386
Db 616 KGALETEDPALIYDYAIVLFKGGVKKNRRLALELMKKKASKGLHOAVNGLGWYHKFK- 674
QY 387 VPLNYAEALYFOKAAEKGWPDQFOLGFMYYSG--SGI-WKDYKLAIFYLASOGOP 443
Db 675 --KNYAKAAKYLWKAEMGNPDASYNLGLDGIFFGVPGRNOTLAGEVFHKAAGQGHM 732
QY 444 LAIYIYAKMYATGT--GVVRSRTAVELYKGYCE-----LGHWAKEKLTAYFAYKGDGIDS 497
Db 733 EGTFLWCSLYITIGNFTFPDPEKAVYWAHVAEKNGYLGHVIRKGLNAYL-----EGSWHE 789
QY 498 SLVQYALLAEMGYEVAQSNFASFTLESKANILEKEKMYPMALLW--NRAAIQNGNA--F 552
Db 790 ALLYVYLAFTGLEVSQTNLAHICEERP----DLARRYLGVCVWRYNFSFQIDAPSF 845
QY 553 ARVKIGDYHYGY-GTKDYQTAATHYSIAANKYHNAQAMFNLAHYEHGLGI----- 604
Db 846 AVLKMGDLYYGHQNSQDLELSVQMYAQALD-GDSQGFENLALLIEGTIIPHHILD 904
QY 605 -----TKDHLARRLYDMA-AQTSPDAHIPVLFAVMKLETHLLRDIL 646
Db 905 LEIDSTLHNSNISILQELYERCSWHSNESFSPCSLAWLYLHLRLWLWAIL 955

RESULT 10
O9XDH8 PRELIMINARY; PRT; 1201 AA.
AC O9XDH8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENHANCED ENTRY PROTEIN ENHC.
GN ENHC.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=130B;
RA Cirillo S.L.G., Gupta M., Lum J., Cirillo J.D.;
RT "Isolation of Novel Loci Involved in Entry by Legionella
RT pneumophila."
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF057704; AAD41587.1; -.
DR InterPro; IPR001440; TPR.
SQ SEQUENCE 1201 AA; 134365 MW; 5C4ABEE5F3526DC4 CRC64;
```

Query Match 12.2%; Score 435; DB 2; Length 1201;  
Best Local Similarity 24.2%; Pred. No. 1.1e-19;  
Matches 183; Conservative 102; Mismatches 246; Indels 224; Gaps 26;

```
QY 34 NYVTQVSVNEIKOYLSHILEQRTSSNVINKRENILKKNQKIRIKGTQNKDKLRNKN 93
Db 498 NQATQFEICQLFYQIGLQMDDASAIIF--YENAAEQKHLGAEYNL-GTL---YLKRGKD 551
QY 94 HLOKQAEKNFTDEGDOLF-----MGIKVLQSS-----KSQKQKEAYLLFAK 136
Db 552 ENDYQOALNWL--DSAPKGNKRAQVVLARILRQGVDPGKEYIKANDEQAMAMLYLSA 609
QY 137 AADMGNLK-----AMKMDALL--FGNF- 158
Db 610 ANDYGAPEYELAEYLARDYNNGLSVDRKQKIALIRKLYOGAVKNKVAEALLPLAFYNAM 669
QY 159 --GVONITAAIQLESRAKESCKAQNALGFLSSYGIGMEYDQAKALIYTFGSAGNM 216
Db 670 DNRKQEQAFKVAEEQAEETGNENALLGLMGLYDRGITADPAKAMYWQ--QAGQNPV 727
QY 217 SQMILGYRLSYGNVLQNCVALSYKKVADYIADTFEKSEG--VPVEKRLTLRPERN 789
```

Db	141	RETSRSKSLUHHHPAAAGGNQSKMALAFYRLR-----QNMVDKAVELYAEIAETAVNSF	193
Qy	254	EKSEGVVP-EKVRL---TERPENLSSNSEILDMDIYQYKFLAERGDVQVQVSLGQLHLI	309
Db	196	LISKDSMAEPVRIHIGTEENKDALRKSERGEDEDFQILEYQAEKGSVAMHKIGLFYYP	255
Qy	310	GRKGLDODYYKALHYFLUKAAKAGSANAMAFIGKMYLEGNAVQNNAATAFKYFSMAASKG	369
Db	256	GLRGLRRDHAKALWFSKAE-----	275
Qy	370	NATGLHCLGLLPHGKGV-PLNYAEALKYFQKAAEKGWPDPAQFQLGFMYYSGSIWKDYK	428
Db	276	-----FNLGGLYYKYGVDKRRNTKAREYFEMAANNEDSGHYNLGLVLYLKGTVKDKDR	331
Qy	429	LAFKYFYLASQSGOPLAIYLAQMAYATGTCVWRSCRTAVELYKGVGCELGHWAEKFLTAYF	488
Db	332	HATKYFEVAANAGOPKAFYQLAKMFHTGVGLTKNLEWATTFYKLVAERGPW-----	382
Qy	489	AYKDGDISLVOYALLA 506	
Db	383	-----SSLRSWALEA 392	
RESULT 12			
O77341			
ID	077341	PRELIMINARY; PRT; 807 AA.	
AC	O77341;		
DT	01-NOV-1998 (TrEMBLrel. 08, Created)		
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)		
DE	PF00550W PROTEIN.		
GN	PF00550W.		
OS	Escherichia falciparum (isolate 3D7).		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=36329;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=3D7;		
RA	Hamlin N., Lawson D., Barrell B.;		
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBDJ databases.		
DR	EMBL; AL008970; CAA15596.1; --		
SQ	SEQUENCE 807 AA; 96686 MW; BA0AE118DF01ABA5 CRC64;		
Query Match 11.0%; Score 391; DB 5; Length 807;			
Best Local Similarity 22.1%; Pred. No. 4 4e-17;			
Matches 152; Conservative 106; Mismatches 237; Indels 194; Gaps			
Qy	24	EHNKROKERNVTQVSVNEIK--QYLSHILEQRTSSNVINKRENLLKKNQKRIKRG 81	
Db	92	EEEEKKNKLYLKNDRANDLDSFSYIHNNEDSSNDNINGKYYNNRKQNDNNKGD 151	
Qy	82	IQNKDIL--KRN-----	
Db	152	YNNKGDYNNKRNYSKNDKDEYNKYTYDNDNNYNDENYNDYNEYNNEIKKEKEK 211	
Qy	104	TDEGDQLFKMGIKVLOQSKOK-EEAYLLFAKADMGNLK---AMEKMAALLFG-- 156	
Db	212	ESDRKQMFKAIELNGSKSKKKNNINKCIEFQKLITDKNDKTRSSYELGKIYFPGYK 271	
Qy	157	NGFY---QNTAAIQLYESLAEGSCKAQNALGFLSSYGI-----	
Db	272	NYPSYKRNVLNLSHYLQKAAMKNPAALHFLSFIYDFPHKTEENKRKEKVIQNQN 331	
Qy	194	-----GHEYDO-----AKALIYTFGSAGNMMSQILGYRYSGLINVLQCEVAL 239	
Db	332	VNNINNNYQNLOQNEFNFKKSEIEFEMIAASLNYPISLTLAYKFLYGINMKQNCYKAK 391	
Qy	240	SYKKVA-----DYIADTFEKGVPVEKVRTERPENLSSNSEILDMDIYQYKFLAE 293	
Db	392	KLYKNVAENYNSDYI-----NPLSEDLN-NGENLNHNEI-----	
Qy	294	RGDVQIQVSLGQLHLIGRKGDDIYKALHYFLKAAKAGSANAMAFITGKMYLEGNAVPO 353	

```

Db 429 -----NMKNNEEIELEFLNEQIKGQDVMAWYDLGKKYKE-----BK 465
QY 354 NNATARKYFSMAASKGNATIGLGLYFGKGVPLNYAEALKYFOKAAEKGWPDQAQFOL 413
Db 466 NFQOAKFYINEASKNNLALALGLIYLYGYGVQKDKINSFNSKAAEAGDVESKCYL 525
QY 414 GMYSSGSGIKWDYKIAFYK-----FYLA-----437
Db 526 GYIYFEIDG-YKNLELSKLYLEAASHDYGEAFFFLAEIILDISMRKQVYISDYVYEVVFK 584
QY 438 -----SQSQOPLAIYVLAKMYATGTVWRSCTAVELYKGVCELGHWAEEKFLTAIPAYKD 492
Db 585 LYEHSAADLGQVQAFREAOQLEYIGKGVKQSCNLATLSYKFAIEASTLWINNRQGM DYLE 644
QY 493 GQIDSSLVQYALAEAGYEVASQNSAFILESKKA-NILEKEKMPWALLLWNRRAATQGNA 551
Db 645 KYDKAFYFYALASYEGYEIAQNNLVYIYRTNKLNNVHPRKI-----MLVNLLYKQGNV 700
QY 552 FARVKIGDYH-----YYGYGTTK 569
Db 701 KALYENGIEYKEQNEELSVSYYKLGKK 729

RESULT 13
ID Q05787 PRELIMINARY; PRT; 833 AA.
AC Q05787;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CHROMOSOME XII COSMID 8167.
GN HRD3 OR L8167.5 OR YLR207W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
RA Entlan K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Klose K., Kottler P.,
RA Louis E.J., Messenguy F., Mewes H.W., Miosga T., Mostl D.,
RA Muller-Auer S., Wentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Pauley A.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Waterston R.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Cherry J.M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U14913; AAB67427.1; -.
DR SGD; S0004197; HRD3.
SQ SEQUENCE 833 AA; 95480 MW; E7DB29FB16D9BFD CRC64;

```

```

Query Match 10.4%; Score 373; DB 3; Length 833;
Best Local Similarity 23.9%; Pred. No. 6.7e-16;
Matches 165; Conservative 106; Mismatches 250; Indels 168; Gaps 28;

QY 97 KOAKNF-----TDEGQLFKGIKIVLQOSKSKQKEEAYLL-----FAKAA 138
Db 68 RNEKNYQSIWONEITDSORHIEL---LVQSEQFNNEATYTLQIHLWSQYNFPHNM 124
QY 139 DMGNLKAMEKMDALLFGNFGVQNTAAIQLYESLAKESGCAQNALGFLSSVIGMEYD 198
Db 125 TLAH-KYLEKFNLDLTHFTNH-----SAIFDLAVMTATGCCASGNDQTVIPO-----D 170
QY 199 QAKALIYTYFGSAGNMMSQMLGYRLSGINVLQCEVALSYKVKVADYIADTFEKS- 257
Db 171 SAKALLYQRAAQLGNLAKAKQVLAYKYSGFVPRNFHSLVLYRDIAEQLRKSYSRDEM 230
QY 258 -----GVPEKVR-LTERPENLSSNSEILWDIYQYK 289
Db 231 DIVFPWESYNNVRISDFESGLLGKGLNSVPSSTVRKTRTRPDIGS-----P 276
QY 290 FLAERGDVQIOVS---LGQHLIGRKGL-----DOD-----YYKALHYELKAAKA 331
Db 277 FIAQVNGVOMTLOIEPMGRFAFNGDNGINGDEDEDASERRIIRIYYAALNDY-KGTYS 335
QY 332 GSAN-----AMAFIGKMYLEGNAAPQNNATAF 359
Db 336 QSRNCERAKNLELTYKEFOPHVDNLDPLQVYVYVRCLOLLGHMYFTGESSKPNHMAE 395
QY 360 KYF--SMAASKNATGLHG-----LGLLYFHGKGPLNYAEALKYFQKA-----AEKGWPD 408
Db 396 EILTTSLEISR-RAQGPICRACIDLGLI---NQYITNNISQISAYYMKANKTQANN--I 449
QY 409 AQOLFQMYSSGSGIKWDYKLA--PKYFVLASQSGOPLAIYVLAKMYATGTVVRSCTA 466
Db 450 VEFQSKLATS---FPEEKIGDPFNLMTAYLNGFIPAIYEFVAMIESGMKSKEVENT 505
QY 467 VELYKGVCELGH--WAEKELTAYFAYKDGIDIDSSLVQYALLAEQMEYVAQSAFTLESK 524
Db 506 AYLKFTFVDKNEAIMAPKLRTAFALINDRSEVALWAYSQLAEQGVETAQVSAAYLMYOL 565
QY 525 KANILE-----KEKMPMALLNWRRAIQGNAPFARKYIGDYHYGYGTTKDYQTAATHYSI 580
Db 566 PYEFEDPPRTDQRTKLTASYTRAFCQGNIDAGVAGDIYF---OMQNYSKAMALYQG 621
QY 581 AANKYHNAQAMENLAYMYEHGLITKDILHARRLYDMAQTSDDAHIPVLFAVMKLETH 640
Db 622 AALKY-SIQAIWNLTGMHEHGLGVNRDHFHARRYDQVSEHDFHFLASKLSVYKLD---H 677
QY 641 LLRDILFFNQFTRWNLKLDNTIGPHWD 669
Db 678 LKSWLTWITR--EKVNYMKPSSPLNPND 704

RESULT 14
Q9NPY6 PRELIMINARY; PRT; 71 AA.
AC Q9NPY6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE DJ631M13.1 (SIMILAR TO MOUSE SEL1L (SEL-1 (SUPPRESSOR OF LIN-12, C.ELEGANS)-LIKE) (FRAGMENT)).
DE C.ELEGANS)-LIKE) (FRAGMENT).
GN DJ631M13.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Laird G.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117333; CAC01943.1; -.

```

FT NON\_TER 71 71  
SQ SEQUENCE 71 AA; 7756 MW; 3EF2C63469879DF4 CRC64;

Query Match  
Best Local Similarity 9.8%; Score 349; DB 4; Length 71;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 113 MGKIVLQSKSQKQKEEAYLLFAKADMGNLKAMEKMDALLFGNGVQNTAAIQLYES 172  
Db 1 MGKIVLQSKSQKQKEEAYLLFAKADMGNLKAMEKMDALLFGNGVQNTAAIQLYES 60  
QY 173 LAKEGSCRAQN 183  
Db 61 LAKEGSCRAQN 71

RESULT 15  
Q9RN76 PRELIMINARY; PRT; 469 AA.  
AC Q9RN76;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
DE IMMUNOREACTIVE PROTEIN.  
OS Coccidia burnetii.  
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;  
OC Coccidia group; Coccidia.  
OX NCBI\_TaxID=777;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NINE MILE PHASE I;  
RA Zhang G.Q., Hota A., Yamaguchi T., Fukushi T., Hirai K.;  
RT "Cloning and characterization of the Coccidia burnetii gene encoding  
an immunoreactive protein."  
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF185288; AAF01236.1; --  
SQ SEQUENCE 469 AA; 52695 MW; DDB341C820011D5 CRC64;

Query Match  
Best Local Similarity 8.2%; Score 294.5; DB 2; Length 469;  
Matches 132; Conservative 86; Mismatches 190; Indels 135; Gaps 22;  
QY 47 YLSHILEQRTSSNVINKRENLEKKKQKIRIKIQNKDI-----LKRKNHQLQKQAE 100  
Db 21 YVSGFL--NPDNNVSLETQNNKKAYALQKAAKSGLEKADVALGYFLQQNTTL--AE 75  
QY 101 KNFTDEGDLFKMGIKVLQSKSQKQKEEAYLLFAKADMG-----NLKAMEKMD 151  
Db 76 KTFE-----IAQRAYQTNDFEAMLLAILYDRGFGVNRNRSKSAEILEKLSK 122  
QY 152 -----ALLFGNGVON---ITAAIQLYESLAKEGSCRAQNALGFLSSVIGIMEYDOAKA 202  
Db 123 QNNATAQFMGLNYLKNRKENIAISLSKANSQNGYAKYNLAILAK-----QNK- 173  
QY 203 LIYYTFGSAGNMMSOMILGYRYLSGINVLQNCVALSYRKVKVADYIADTFEKSEGVPE 262  
Db 174 ---YT--KPGENFLSLLIR-----AANHYDKIELADYLLDTPVP-- 210  
QY 263 KVLRLTERPENLSSNEILDYIYQYKFLAERGVOIOVSLGOLHLIGRKGL--DQDYK 320  
Db 211 -----GSEKAVAIQF---LANKQDPAAELKLG---FMNEHGLLFPKDYHK 251  
QY 321 ALHYFLKAAGKAGSANAMAFICKMYLEGNAVPPNNAATAFKYFSMAASKGNAIGLHGLL 380  
Db 252 AEEYOKSAEOGNPIAQYLLGNMYLGR--GVDRVKNKAIDMLKKSAAQNYVPKVGGLGFI 310  
QY 381 YPHGKGPLNTVAALKYFQKAAKGPDAQFOLGPGMYSGGINKDYKLAIFYFYLASOS 440  
Db 311 YEMSKH---NYPEAKKWTLSKFPNPOALYNLGLMYEYKGVKSDPKAFRLYKDAQN 367  
QY 441 GQPLAIYYLAKMYATGTGVVRSCTRTAVELYKGVCELGHWAEEKFLTAYFAYKDGIDIDSLV 500

Db 368 GLDLAAVQVAGMYLKGTTGIGFDPTALRMYSQAAQKN---SFAT-----Y 410  
QY 501 QYALLAEMGYEVAQSNLSAFILSKKANILEKEKMYPMALLLNRAAIOGNAFARVKIGDY 560  
Db 411 QLGLMSESG--VAQKID-----LNKARLY-----YEKAKEGSGVEAQLALARF 451  
QY 561 HYY 563  
Db 452 YEF 454

Search completed: April 8, 2002, 15:08:14  
Job time: 228 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2002, 20:35:13 ; Search time 1542.33 Seconds  
(without alignments)  
22141.281 Million cell updates/sec

Title: US-09-714-882-1  
Perfect score: 2070  
Sequence: 1 atgaagccctgtctctgtt.....ttagaatcaccatgggtag 2070

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo\_hum.\*
- 31: em.htgo\_inv.\*
- 32: em.htgo\_rod.\*
- 33: em.htg\_hum.\*
- 34: em.htg\_inv.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	2069.6	100.0	2070	6	AX148133	AX148133 Sequence
2	2053.6	99.2	2067	6	AX148135	AX148135 Sequence
3	1650.2	79.7	1773	6	AX148137	AX148137 Sequence
4	1593.8	77.0	2129	9	HSN802045	AL137678 Homo sapi
5	1403.6	67.8	1500	6	AX148141	AX148141 Sequence
6	1403.6	67.8	1728	6	AX148145	AX148145 Sequence
7	1403.6	67.8	1731	6	AX148143	AX148143 Sequence
8	1254.6	60.6	1257	6	AX148139	AX148139 Sequence
9	631.4	30.5	2382	6	E26632	E26632 TSA305 gene
10	631.4	30.5	3473	9	AF052059	AF052059 Homo sapi
11	631.4	30.5	7885	9	E26633	E26633 TSA305 gene
12	612.2	29.6	2441	10	AB048195	AB048195 Mesocric
13	612.2	29.6	2441	10	AB048195	AB048195 Mesocric
14	607.2	29.3	3663	10	AF063095	AF063095 Mus muscu
15	510.4	24.7	557	6	AX079694	AX079694 Sequence
16	444	21.4	2144	6	AX136309	AX136309 Sequence
17	368.4	17.8	4264	3	AF181652	AF181652 Drosophill
18	331.6	16.0	2406	3	CEU50829	U50829 Caenorhabdi
19	179.6	8.7	100272	9	HSJ842G6	AL109657 Human DNA
20	172.4	8.3	156813	9	HSJ631M13	AL117333 Human DNA
21	172.4	7.9	37434	3	CEP45B3	Z78063 Caenorhabdi
22	163.2	7.9	3613	3	CEU50828	U50828 Caenorhabdi
23	148	7.1	2322	8	AY039982	AY039982 Arabidops
24	143.2	6.9	453	10	AF304855	AF304855 Rattus no
25	117	5.7	1672	6	AX092286	AX092286 Sequence
26	115.4	5.6	109151	2	AC017637	AC017637 Drosophill
27	115.4	5.6	161278	3	AC008223	AC008223 Drosophill
28	115.4	5.6	239327	3	AE003744	AE003744 Drosophill
29	112.2	5.4	190153	3	AC008201	AC008201 Drosophill
30	90.2	4.4	571	6	AR107131	AR107131 Sequence
31	90.2	4.4	571	9	HSU11037	U11037 Homo sapien
32	85.4	4.1	482	9	AB0247AS15	AB024761 Homo sapi
33	85.4	4.1	167339	2	CNS01DSD	AL121769 Homo sapi
34	85.4	4.1	174788	9	CNS01DWU	AL136039 Human chr
35	85.4	4.1	180254	2	AC084001	AC084001 Homo sapi
36	85.4	4.1	182799	2	AC073526	AC073526 Homo sapi
37	83.2	4.0	34001	6	AX067443	AX067443 Sequence
38	82.4	4.0	761	11	CNS05KKG	AL402998 T7 end of
39	81.4	3.9	70511	2	AC091083	AC091083 Homo sapi
40	80.8	3.9	231972	2	AC068055	AC068055 Homo sapi
41	80.6	3.9	156550	2	AC015830	AC015830 Homo sapi
42	80.2	3.9	6644	6	E23356	E23356 Virus vecto
43	80.2	3.9	7218	6	I66494	I66494 Sequence 14
44	80.2	3.9	7372	6	E23357	E23357 Virus vecto
45	80.2	3.9	7797	6	E23355	E23355 Virus vecto

ALIGNMENTS

RESULT 1  
AX148133  
LOCUS AX148133  
DEFINITION Sequence 1 from Patent WO0136636.  
ACCESSION AX148133  
VERSION AX148133.1 GI:14347059  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2070)  
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.  
TITLE Human notch ligand proteins and polynucleotides encoding the same  
JOURNAL Patent: WO 0136636-A 1 25-MAY-2001;  
Lexicon Genetics Incorporated (US)  
FEATURES  
source  
1..2070  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

PAT 08-JUN-2001

BASE COUNT	701	a	379	c	452	g	537	t	1	others
ORIGIN										
Query Match			100.0%;	Score	2069.6;	DB	6;	Length	2070;	
Best Local Similarity			100.0%;	Pred. No.	0;					
Matches 2070;			Conservative	0;	Mismatches	0;	Indels	0;	Gaps	
Qy	1	atgaagccctgtctctgttaatagagatatatgataatcttggggtgcacaaataaaact	60							
Db	1	ATGAAGCCCTGTCTCTGTTAATAAGAGATATGTGATATTTCTTGGGTGCATATTAACACT	60							
Qy	61	atcaaacagaggaacataataaaagacaaaaggaaagaatgtcaccacacaggtatca	120							
Db	61	ATCAAGCAGAGGAACATAATAAAGACAAAAGGAAAGAAATGTCACCACACAGGTATCA	120							
Qy	121	gtgaacgaatcaaacatattttatcacacatatgtgaaacaaagaacatctagttaatgta	180							
Db	121	GTGAACGAATTCAAACAATATTTATCACACATATTGGAACAAGAACAATCTAGTAATGTA	180							
Qy	181	atcaataaagagagaaatccctgggagaaaaagagaaatcaacgtlaagataagaataaaa	240							
Db	181	ATCAATTAAGAGAAATCTCTCGGAGAAAAGAGAATCAACGTAAATTAAGATAAAA	240							
Qy	241	ggaattcaaaataagatatctttgaagagaaataaagaaatcatttcaaaagcaagcagag	300							
Db	241	GGAAATCAAAATAAAGATATCTTTGAAGAGAAATAAGAATCATTTACAAAGCAAGCAGAG	300							
Qy	301	aaaaattttcacagatgaagagaccagctatttaagtatgggcatcaaggtttctccagcag	360							
Db	301	AAAAATTTTACAGATGAAGGAGACCAGCTATTTTAAGATGGGANTCAAGGTTCTCCAGCAG	360							
Qy	361	tctaaagccaaaaacaaaagaaagagcctaccctactttttgccaagagcgtgacatg	420							
Db	361	TCTAAAGCCAAAACAAAAGAAAGAGGCTACCTACTTTTGGCCAAAGCAGCTGACATG	420							
Qy	421	ggaaacttgaagctatggagaaaaatggctgacgctttgtctatttggaaatttggggtg	480							
Db	421	GGAAACTTGAAGCTATGGAGAAAATGGCTGACGCTTTGCTATTTTGGAAATTTGGCGTG	480							
Qy	481	caaaataacagcagctatccaattatatatgatgctcttggctaaagaaggtatcatgtaaa	540							
Db	481	CAAAATATAACAGCAGCTATCCAAATTTATATGAGTCTCTGGCTTAAAGAAGGATCATGTAAA	540							
Qy	541	gcccaaacgcattaggatttttgcctctctatggaataggaatggaatatgatcaagct	600							
Db	541	GCCCAAAACGCATTAGGATTTTGTCTCTTATGGAATAGGAATGAATATGATCAAGCT	600							
Qy	601	aaggcactgatattacacctttggaagtgcgtgagaaacatgatccacagatgatt	660							
Db	601	AAGGCACTGATATATTACACTTTGGAAGTGTGGAGAAACATCATGCTCCAGATGATT	660							
Qy	661	tgggggtacagatatatttgcgggaatcaatgtctacagaaatgtgaagttgcctcaagt	720							
Db	661	TTGGGGTACAGATATTGTGCGGAATCAATGTCTACAGAAATTTGTGAAGTTTGCCTAAGT	720							
Qy	721	tattacaagaagtggcagattatatgtctgacacatttgaaaaaagtgaaggtgttcca	780							
Db	721	TATTTACAAGAAAGTGCAGATTTATTTGCTGACACTTTGAAAAGGTGAAGGTTTCCCA	780							
Qy	781	gtggaaaagtgcagctaacggaagacacctgaaatctctagttcttaacagtgagattttg	840							
Db	781	GTGGA AAAAGTGAGACTAACGGAAGACCTGAAATCTGAGTTCTTAACAGTGAGATTTTG	840							
Qy	841	gattggacatataccaatactataaatttttggcagaaaggagatgtttcagatacaa	900							
Db	841	GATTGGACATATACCAATACTATAAATTTTGGCAGAAAAGAGAGATGTTTCAGATACAA	900							
Qy	901	gtctctcttgacaattacatctaattggcaggaagaagctctcagatcaggattactacaaa	960							
Db	901	GTCCTCTTGCACAATTAATCATCTTAATTGGCAGGAAGGCTAGATCAGGATTACTACAAA	960							
Qy	961	gcattacactactctttaaggcagcaaggccggagtgcaaatgccatgacatttata	1020							

Db 2041 ATTTTGGTTCITAGAAATCACCATGGGTAG 2070

RESULT 2  
AX148135 2067 bp DNA PAT 08-JUN-2001  
LOCUS  
DEFINITION Sequence 3 from Patent WO0136636.  
ACCESSION AX148135  
VERSION AX148135.1 GI:14347060  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2067)  
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.  
TITLE Human notch ligand proteins and polynucleotides encoding the same  
JOURNAL Patent: WO 0136636-A 3 25-MAY-2001;  
Lexicon Genetics Incorporated (US)  
FEATURES  
source location/Qualifiers  
1..2067  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 700 a 378 c 451 g 537 t 1 others  
ORIGIN

Query Match 99.2%; Score 2053.6; DB 6; Length 2067;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2067; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 atgaagccctgtctctgttaagagatatgataattcttgggtgcacataaaact 60  
Db 1 ATGAAGCCCTTCTCTCTGTTAATAGAGATATTGATAATTTCTTGGGTCAACAATTAATAACT 60  
QY 61 atcaagcagagaaacataataaagacaaagaaagaaatgtcacacacaggtatca 120  
Db 61 ATCAAGCAGAGAAACATAATAAAGACAAAGAAAGAAATGTCAACACAGGTATCA 120  
QY 121 gtgaacgaatacaacaattattcacacatatgtggaacaaagaacatctagtaatga 180  
Db 121 GTGAACGAATCAACAATATTATACACACATATTGGACAAAGAACATCTAGTAATGTA 180  
QY 181 atcaataaaagaaatactctgtgagaaaaagaagaatcaacgtaaaaataagaataaaa 240  
Db 181 ATCAATAAAAGAAAAATCTCTGTGAGAAAAAGAAAGAAATCAACGTAAAAATAAGATAAAA 240  
QY 241 ggaattcaaaataagagatatcttgagagaaataagaatcatttcaaaagaacagag 300  
Db 241 GGAATTCAAAATAAGAGATATCTTGAGAGAAATAAGAAATCATTTACAAAGCAAGCAGAG 300  
QY 301 aaaaatttcacagatgaaggagaccagctatttaagatgggcatcaaggttctccagcag 360  
Db 301 AAAAATTTACAGATCAAGGACACAGCTATTATAGATGGGCATCAAGGTTCTCCAGCAG 360  
QY 361 tctaaagcacaacaaagaagaagcctacctactcttttgcacaaagcagctgacatg 420  
Db 361 TCTAAAAGCAAAACAAAAGAAAGAGCTTACCTACTTTTGCACAAAGCAGCTGACATG 420  
QY 421 ggaacttgaagctatggagaaaatggctgcgctttgctatttggaaaatttggcgcg 480  
Db 421 GGAAACTTGAAGACTATGGAGAAAAATGGCTGACGCTTTGCTATTTTGGAAATTTGGCGCTG 480  
QY 481 caaataaacagcagctatcaattatatagctgcttgctgctaaagaagcatcatgtaa 540  
Db 481 CAANATATAACAGAGCTATCCANATATATAGTCTCTGGCTTAAGAGAGATCATGTAAA 540  
QY 541 gcccaaacgcattaggattttgtctcttattggaatagggaatggaatatgataagct 600  
Db 541 GCCCAAAACGCATTAGGATTTTGTCTTCTTATGGAATAGGAATGGAATATGATCAAGCT 600  
QY 601 aaggcactgatattacacaccttggagtgctggaggaaacatgatgtcccagatgatt 660  
Db 601 ATTTTGGTTCITAGAAATCACCATGGGTAG 2070

Db 601 AAGGCACGTGATATATATACACCTTTTGAAGTGTGGAGGAAACATGATGTCCCAGATGATT 560  
QY 661 ttgggttacagatatattgtcgggaatcaatgttctacagaatttgaagtccctaaagt 720  
Db 661 TTGGGTACAGATATTGTGCGGAATCAATGTTCTACAGAAATTGTGAAGTTGCCCTAACT 720  
QY 721 tattacaagaagtggcagattatatgtctgacacatttgaaaaaagtgaaaggtttcca 780  
Db 721 TATTACAAGAAAGTGGCAGATTATATTGCTGACACATTTGAAATAAGTGAAGGTGTCCA 780  
QY 781 gtgaaaaagtgaactaacgaaagacctgaaaactctgagttctaaacagtgcagatttg 840  
Db 781 GTGGAANAAGTACAGACTAACGGAAGACCTGAAATCTGAGTTCTAACAGTGCAGATTTG 840  
QY 841 gattgggacataacaaataactataaattttggcagaagaaggagatgttcagatacaa 900  
Db 841 GATTGGGACATATACCAATFACTATAAATTTTGGCAGAAAGAGGAGATGTTCAGATACAA 900  
QY 901 gtctcttggacaaattacatttaattggcagaaaggtctagatcaggattactacaaa 960  
Db 901 GTCTCTCTTTGGACAATTACATCTAATTGGCAGAAAGGTCTAGATCAGGATTACTACAAA 960  
QY 961 gcattacactactctttaaagcagcaaaagcgcgggagtgcaaatcccatggcattata 1020  
Db 961 GCATTACACTACTTCTTAAAGCGACCAAGGCGGGAGTGCATAATGCCATGGCATTTATA 1020  
QY 1021 ggaagaatgtatttagaggggaatgctgcgtgcgcgcaaaaaaacgactactgccttcag 1080  
Db 1021 GCAAGATGATATTAGAGGGGAATGCTGCGTGCSCCAAAATAACGCTACTGCTTCAAG 1080  
QY 1081 tactttccatggcgcagtgtaaggcgaatgcaatggccttcagtgggttgggttctctt 1140  
Db 1081 TACTTTTCCATGGCAGCCAGTAGGAGCAATGAATCGGCTTCATGGGCTGGTCTCTCTT 1140  
QY 1141 tactttcatggaaaaaggadtccctcgaattatgccraagcacttaaaactcttcagaaa 1200  
Db 1141 TACTTTTCTGGAAGAGGAGTTCCTCTGAAATATTGCCRAAGCACTTAAATACITTCAGAAA 1200  
QY 1201 gctgcgaaaaaggggtggcccgacgcacagttccagttcagttcagttcagttcagtt 1260  
Db 1201 GCTGCGAAAAAGGTTGGCCGACGCACAGTTCCAGTTAGGCTTTCATGTACTCTGGC 1260  
QY 1261 tctggaatatgaagattataaaacttgcttcaaatatattttacccctggcatctcagagt 1320  
Db 1261 TCTGGAATATGAAGGATTATAAACTTGCTTCCAAATATTTTACCTGGCATCTCAGAGT 1320  
QY 1321 ggagccctccctgcatttattctggcagaatgtatgcacaagagaaagagtagta 1380  
Db 1321 GGGCAGCCCCCTGCCATTTATTATCTGGCCAAAGATGATGCAACAGAGAGAGTAGTA 1380  
QY 1381 agatcatgcagaactgctgtgagctttataaagggtgtctgtgaactagccactgggct 1440  
Db 1381 AGATCATGCAGAACTGCTGTGGAGCTTTATAAGGTGTCTGTGAACCTAGGCCACTGGGCT 1440  
QY 1441 gagaaattcctgacagcttactttgctataaggatgggtgatagattctctctgtt 1500  
Db 1441 GAGAAATTCCTGACAGCTTACTTTTGCTATAGAGTGGTGATATAGATTTCTCTCTGT 1500  
QY 1501 cagtatgcactgtgcagaaatggggtatgaatagtagtcaaaagaattcagcattcatt 1560  
Db 1501 CAGTATGCACTGCTGCAAAAATGGGTATGAAGTAGTACGTCAAAGCAATTTTTCAGCATTT 1560  
QY 1561 ttggaattcaaaaggtcaacattcttgaagaagaagatgtatcaaatggcgctctc 1620  
Db 1561 TTGGAATCTAAAAGGCTAACATTTTGAANAAGAGAAAGATGATATCAATGGGCTCTC 1620  
QY 1621 ctatggaattcgagctgccatttcaaggcaatgcatttgcctagagtaaaaaattggagattac 1680  
Db 1621 CTATGGAATCGAGCTGCCATTCAGGCAATGCAATTTGCTAGAGTAAAAAATTTGGAGATTAC 1680  
QY 1681 cactactatggctatgggactaagaagaactatcaaaagcagcagcacaactacagcatt 1740  
Db 1681 CATTACTATGGTATGGGACTAAGAAAGACTATCAAAACAGCAGCCACACTACAGCATTT 1740

```
QY 1741 gcagccaaacataccacaacgcgcaagccatgttcaactctgcttatgtatgaacac 1800
|||||
Db 1741 GCAGCCAAACATAACACAAACGCGCAAGCCATGTTCAATCTGGCTTATATGATGACAC 1800
|||||
QY 1801 ggttaggcatacaaaagacattccacttgccgcagaagattgtacgacatggctgctcaa 1860
|||||
Db 1801 GGTTAGGCATACAAAGACATTCACCTGGCCAGAAATTTGACACATGGCTGCTCAA 1860
|||||
QY 1861 acgagtcagatgccacataactgtctcttgcgcgtcatgaaactggaactacgcgat 1920
|||||
Db 1861 ACGAGTCCAGATGCCACATACCTGTGCTTTGGCCGTCATGAAACTGGAAACTACGCAT 1920
|||||
QY 1921 ttgctcgggatatctctgtttttaaactagttccacaacagatggaactggctgaaactg 1980
|||||
Db 1921 TTGCTCGGGATATCTGTGTTTTTAAT---TTCACAACGAGATGGAACCTGGCTGAAACTG 1977
|||||
QY 1981 gacaacaccattggaccacactgggacttattgtgattggcctcattgttctcgggctg 2040
|||||
Db 1978 GACAACACCATTGGACCACACTGGGACTTATTTGTGATTGGCTCATTTGCTGGGCTG 2037
|||||
QY 2041 atttgttcttagaataccactgggtag 2070
|||||
Db 2038 ATTGTTGCTGTAGAAATCACCATGGGTAG 2067
|||||

RESULT 3
AX148137
LOCUS AX148137 1773 bp DNA PAT 08-JUN-2001
DEFINITION Sequence 5 from Patent WO0136636.
ACCESSION AX148137
VERSION AX148137.1 GI:14347061
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1773)
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.
TITLE Human notch ligand proteins and polynucleotides encoding the same
JOURNAL Patent: WO 0136636-A 5 25-MAY-2001;
Lexicon Genetics Incorporated (US).
FEATURES
source
1. .1773
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 607 a 313 c 388 g 464 t 1 others
ORIGIN
Query Match 79.7%; Score 1650.2; DB 6; Length 1773;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1653; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgaagcccttctctgttaagatattgataattctctggggtcacaaattaaaact 60
|||||
Db 1 ATGAAGCCCTTCTCTGTGTTAAGATATTGATAATTCTTGGGCTCACAAATTAATACT 60
|||||
QY 61 atcaagcagaggaacataataaagacaaaaggaagaaatgtccacacacaggtatca 120
|||||
Db 61 ATCAAGCAGAGACATAATAAAGACAAAGGAAGAAATGTCCACACACAGGTATCA 120
|||||
QY 121 gtgaacgaatacaacaattatttatcacacatttggaacaaagaacatctagtaatga 180
|||||
Db 121 GTGAACGAATCAACAATATTATCACACATATTGGAACAAAGAACATCTAGTAATGA 180
|||||
QY 181 atcaataaagagaataactctcgtgagaaaaagaagaatcaacgctaaataagaataaaa 240
|||||
Db 181 ATCAATAAAGAGAAATCTCTCTGAGAAAAAGAAAGATCAACGCTAAATAAGAAATAAA 240
|||||
QY 241 ggaattcaaaataaagatatcttgagagaaataagaatacatttcaaaaagcaagcagag 300
|||||
Db 241 GGAATTCAAAATAAGATATCTTGAAGAGAAATAGATATCAATTTTCAAAAAGCAAGCAGAG 300
|||||
```

```
QY 301 aaaaattttacagatgaaggaacacagctatttaagatgggcatcaagttctccagcag 360
|||||
Db 301 AAAAATTTTACAGATGAAGGAGACAGCTATTATAGATGGCATCAAGTTTCTCCAGCAG 360
|||||
QY 361 tctaaaagcaaaaacaaaagaagcctacctactcttttggccaaagcagctgacatg 420
|||||
Db 361 TCTAAAAGCCAAAACAAAACAAAGAGAGCTACTACTTTTTTGCCTAAACAGCAGCTGACATG 420
|||||
QY 421 ggaacttgaaagctatggagaaaaatggctgacgccttctgctatttggaaaaatttggcgtg 480
|||||
Db 421 GGAACACTTTGAAAGCTATGGAGAAAATGGCTGACGCTTTGCTATTATTTGGAAATTTTGGCGTG 480
|||||
QY 481 caaaataaaacagcagctatccaattatagctcttgctgctaaagaagatcaatgtaaa 540
|||||
Db 481 CAAAATATAACAGCAGCTATCCAAATATATGAGTCTCTGGCTAAGAGAGATCATGTAA 540
|||||
QY 541 gcccaaacgcattaggatttttctcttcttgaataggaaatggaatgatcatcaagct 600
|||||
Db 541 GCCCAAAACGCATTAGGATTTTGTCTTCTTATGGAATAGGAATGGAATATGATCAAGCT 600
|||||
QY 601 aaggcactgatatatacacaccttggaaagtgctggaggaacacatgatgtcccagatgatt 660
|||||
Db 601 AAGGCACATGATATATACACCTTTTGAAGTGTGGAGGAAACATGATGTCCAGATGATT 660
|||||
QY 661 ttggggtacagatatattgtcgggaatcaatgttctacagaattgtgaagtgcctcaagt 720
|||||
Db 661 TTGGGTACAGATATTGTTCGGGAATCAATGTCTACAGAAATGTGAAGTTCGCCCTAAGT 720
|||||
QY 721 tattcaagaagtggcagatttatattgctgcacacatttgaaaaaagtgaaagtggttcca 780
|||||
Db 721 TATTACAAGAAAGTGGCAGATTATATTGCTGACACATTGTGAAAAAAGTGAAGGTGTCCA 780
|||||
QY 781 gtgaaaaagtgagactaacggaaagacctgaaatctgagttctaaacagtgcagatttg 840
|||||
Db 781 GTGAAAAAGTGAGACTAACGGAAAGACCTGAAAAATCTGAGTCTTAACAGTGCAGATTTTG 840
|||||
QY 841 gattggacatacaactactactataattttggcagaagaggagatttcagatacaaa 900
|||||
Db 841 GATTGGACATATACCAATACTATAAATTTTGGCAGAAAGAGAGATGTTCAGATACAA 900
|||||
QY 901 gtctcttggcaattacatctaattggcagaagaagtcctagatcagagattactacaaa 960
|||||
Db 901 GTCTCTCTTGGCAATTAATCATTAATTTGSCAGAAAGGCTCTAGATCAGGATTACTACAAA 960
|||||
QY 961 gatttacactactcttaaggcagcaaaagccggagtgcaaatgccatgcatttata 1020
|||||
Db 961 GCATTACACTACTCTTTAAGCAGCAAAAGCCGGAGTGCAAATGCCATGCCATTTATA 1020
|||||
QY 1021 ggaagatgtatttagaggggaatgctgcgctgcccgaataaactgcctactgccttcaag 1080
|||||
Db 1021 GGAAGATGTATTTAGAGGGGAATGCTGCCGTGCCGCAAAATAACGCTACTGCTTCAAG 1080
|||||
QY 1081 tactttccatggcgccagtaagggcaatgcaatcgcccttcacgggcttggtgtcttt 1140
|||||
Db 1081 TACTTTTCCATGGCAGCCAGTAAGGCCAATGCAATCGGCCTTTCATGGGCTGTGCTCTT 1140
|||||
QY 1141 tactttcatgaaaagagagttccctgaattatgccraagcacttaatactttcagaaa 1200
|||||
Db 1141 TACTTTTCATGAAAAGAGAGTTCCTCCTGAATATATGCCRAGCACTTAATATCTTCAGAAA 1200
|||||
QY 1201 gctcggaaggggtggccgcagcacagttccagttcaggttaggcttactactctggc 1260
|||||
Db 1201 GCTCGGAAAAAGGCTGGCCCGCACAGATTCACAGTTCCAGTTAGGCTTCTATGTACTCTGCG 1260
|||||
QY 1261 tetggaatatggaagattataaaacttgcccttcaaatatttttacctggcatctcagagt 1320
|||||
Db 1261 TCTGGAATATGGAAGGATTATAAACTTGCCCTTCAATATTTTACTTGCATCTCAGAGT 1320
|||||
QY 1321 gggcagccctcgcatttattctgtgcgaagatgtatgcaacaggaacagagtagta 1380
|||||
Db 1321 GGGCAGCCCTGCCCTATTATTATCTGCCCAACATGTATGCAACAGGAAACAGAGTAGTA 1380
|||||
```

```
QY 1381 agatcatgcagaaactcgtctgtgagatttataaaaggtgtctgtgaaactaggccactgggct 1440
Db 1381 AGATCATGCAGAACTGCTGTGAGGCTTATAAAGGTGTCTGTGAACCTAGGCCACTGGGCT 1440
QY 1441 gagaaattcctgacagcttacttccctataaaggatgggtgatagattcttctctgtt 1500
Db 1441 GAGAAATTCCTGACAGCTTACTTTGCCCTATAGAGTGGTGATATAGATCTTCTCTGT 1500
QY 1501 cagtgactcgttcgcagaaatggggatgaagttagctcctcaagcaattcagcattcatt 1560
Db 1501 CAGTATGCACCTGTCGAGAAATGGGTATGAAGTAGCTCAAGCAAAATTCAGCATTCATT 1560
QY 1561 ttggaatctaaaaggctaaacttcttgaagaaagagagatgtatcccaatgctcttc 1620
Db 1561 TTGGAATCTAAAAGGCTAAACATTTTGAAGAAAGAGAGATGTATCCAAATGGCGTTCTC 1620
QY 1621 ctatggaatcgcagctgcattcaaggcaatgcattg 1657
Db 1621 CTATGGAATCGAGCTGCCATTCAAGGACATTCACTTG 1657

RESULT 4
LOCUS HSM802045 2129 bp mRNA PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp434C1826 (from clone DKFZp434C1826).
ACCESSION AL137678
VERSION AL137678.1 GI:6807900
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2129)
AUTHORS Ansorge,W., Wirkner,U., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-2000) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp434C1826) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES
source
1..2129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp434C1826"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="testis"
polyA_signal 2100..2105
polyA_site 2118
BASE COUNT 713 a 413 c 486 g 517 t
ORIGIN

Query Match 77.0%; Score 1593.8; DB 9; Length 2129;
Best Local Similarity 91.2%; Pred. No. 5.8e-306;
Matches 1839; Conservative 1; Mismatches 3; Indels 174; Gaps 7;

QY 57 aactatcaagcagagggaacataataaaagacaaaaggaagaaatgtcacacacagct 116
Db 194 AGCTATCAAGCAGAGGAACATAATAAAAGACAAAAGGAAGAAATGTCCACCACAGGT 253
QY 117 atcagtgaaacgaatacaacaattattcacacattttggaacaaagacatctagtaa 176
Db 254 ATCAGTGAACGAATCAACAATATTTATCACACATATTGGGAACAAGAACAATCTAGTAA 313
```

```
QY 177 tgtaatcaataaaagagaaaaatctctctggagaaaaagaaatcaacgtataaataagaa 236
Db 314 TGTAAATCAATAAAGAGAAAAATCTCTCTGGAGAAAAAGAAATCAATGTAAAAATAAGAAAT 373
QY 237 aaaaagaattcaaaataaaagatacttgaagagaaataaagaatcatttacaagaacac 296
Db 374 AAAAGGAATTCAAAATAAAGATATCTTGAAGAGAAATAAGAATCATTTTACAAAGCAAGC 433
QY 297 agagaaaaattttacagatgaaggagaccagctattttaagatgggcatacaagttctcca 356
Db 434 ACAGAAAAATTTTACAGATGAAGGAGACCAGCTATTTAAGATGGGCATCAAGGTTCTCCA 493
QY 357 gcagtctaaaagcccaaaaaacaaaaagaagaagcctcactcttbtgccaagcagctga 416
Db 494 GCAGTCTAAAGCCAAAAACAAAAAGAGAACGCTACCTTACTTTTGGCCAAAAGCAGCTGA 553
QY 417 catggaaaacttgaagctatggaaaaatggctacgccttctcttcttggaaatttgg 476
Db 554 CATGGAAAACCTTGAAGCTATGGAGAAAATGGCTGACGCTTTGCTATTTTGGAAATTTGG 613
QY 477 cgtgcaaaataaacagcagctatccaattatatgatgctcttggctaaagaagat-ca 535
Db 614 CGTGCAAAATATAACAGCAGCTATCCAATATATGAGTCTCTTGGCTTAAGAGAGGATCCAT 673
QY 536 gtaaaagcccaaacgcatttaggatttttctctcttcttcttcttcttcttcttcttcttct 595
Db 674 GTAAAGCCCCAAAACGCATTAGGATTTTCTCTCTTCTTATGAGTAGGAAATGGAATATGATC 733
QY 596 aagctaaggcactgatataattacacaccttgggaagtgctggagaaacacatgatgccaga 555
Db 734 AAGCTAAGCAGCTGATATATTACACCTTTTGAAGTGTCTGGAGAAACATGATGTCCAGA 793
QY 656 tgatttggggtcacagatatttgcgggaatcaatgttctcagaaattgtgaagtggccc 715
Db 794 TGATTTTGGGTTACAGATATTTTGTGGGAATCAATGTTCTACAGAAATTTGAAAGTTGCC 853
QY 716 taagtattacaagaagtggcagatttatattgctgacacatttggaaaaagtgaagggtg 775
Db 854 TAAGTTATTACAAGAAAGTGGCAGATTATATTGCTGCACACATTTTGAAGAAAGTGAAGGTG 913
QY 776 ttccagtgaaaaagtgaagtaacggaagacccctgaaactgtgagttcttaacagtgaga 835
Db 914 TTCCAGTGGAAAAAGTGAGACTAACCGGAAAGACCTTGAAATCTGAGTTCTTAACAGTGAGA 973
QY 836 ttttgattggacatataccaataactataaaatttttggcagaaagagagatgttcaga 895
Db 974 TTTTGGATTGGACATATACCAATACTATAAATTTTGGCAGAAAGAGAGATGTTTACA 1033
QY 896 tacaagtc-tctcttggacaattacatctaatgtggcaggaaggtctagatcaggattac 954
Db 1034 TACAAGTCTTCTCTTGGACAATTACATCTAATTGGCAGGAAAGGTCTAGATCAGGATTAC 1093
QY 955 tacaagaattacactactacttcttaaaaggcagcaaaagccgggagtgcaaatgccatggca 1014
Db 1094 TACAAGCATTTACACTACTTCTTAAAGGCGCAAGAGCCGGAGTCAAAATGCCATGGCA 1153
QY 1015 ttatataggaagatgtatttagagggaatgctgcgctgcccacaaataacgctactgcc 1074
Db 1154 TTTATAGGAAGATGTATTTTAGAGGGGAATGCTGCGCTGCCGCAAAATTAACGCTACTGCC 1213
QY 1075 ttcaagtacttttccatggcagccagtaagggaatgaatgcattcggccttcattgggttgg 1134
Db 1214 TTCAAGTACTTTTCCATGGCAGCAGTAAGGGCAATGCAATCGGCCTTCCATGGGCTTGGT 1273
QY 1135 cttcttacttctcatggaaaaagagttcccttgaaattatgcoraagcacttaaaacttt 1194
Db 1274 CTCTTTTACTTTTATGGAAAAAGAGTTTCCCTCTGAATATGCCGAAGCACTTAATACTTT 1333
QY 1195 cagaagctcgcgaaaaaggggtggcccgacgcacagttccagttcaggtcttactgtactac 1254
Db 1334 CAGAAAGCTGCGGAAAAAAGGGT-GCCCGACGCACA-TTCCAGTTAGGCTTCATGTACTAC 1391
```

QY 1255 tctggctctggaatattggaagattataaactgcttcaaatatttttacctggcatct 1314  
 |||||  
 Db 1392 TCTGGCTCTGGAATATGGAAGATTATAAACTTGCCTTCAAAATATTTTACCTGGCATCT 1451  
 |||||  
 QY 1315 cagatgggcagccccctgcccatttattatctctggcgaagatgtatgcaacaggaacagga 1374  
 |||||  
 Db 1452 CAGAGTGGGCAGCCCCCTGCCCATTTTATCTGGCCAAAGATGTATGCAACAGGAACAGGA 1511  
 |||||  
 QY 1375 gtagttaagatcatgcagaactctgtggagctttataaagggtctgtgaaactaggccac 1434  
 |||||  
 Db 1512 CTAGTAAGATCATGCAGAACTGCTGTGGAG----- 1541  
 |||||  
 QY 1435 tgggctgagaaattcctgacagcttacttctgctataaggatgggtgatagattctct 1494  
 |||||  
 Db 1542 ----- 1541  
 |||||  
 QY 1495 cttgttcagtatgcactgctgcagaaatggggtatgaagtagctcaaaagcaattcagca 1554  
 |||||  
 Db 1542 ----- 1541  
 |||||  
 QY 1555 ttcatcttggaattctaaaggcttaacattcttgaagaagagaagatgtatccaatggcg 1614  
 |||||  
 Db 1542 -----AAAGGCTAACATCTTGAAAAAGAGAAGATGTATCCAATGGCG 1585  
 |||||  
 QY 1615 cttctctatggaatcgagctgcccattccaaagcaatgcatttctgtagatgaataattgga 1674  
 |||||  
 Db 1586 CTTCTCTTATGGAATCGAGCTGCCATTCAGGCAATGCTATTTGCTAGAGTAAAAATTGGA 1645  
 |||||  
 QY 1675 gattaccattactatgctatgggactgaagaagactatacaacagcagccacacactac 1734  
 |||||  
 Db 1646 GATTACCATTTACTATGGCTATGGGACTAAGAAGAGACTATCAAACAGCAGCCACACTAC 1705  
 |||||  
 QY 1735 agcattgcagccacaataaccacaacgcgcgaagccatgttccaatgcttcttatgtat 1794  
 |||||  
 Db 1706 AGCAATTGCAGCCAAACAAATACCACACGCGCAAGCCATGTTCAATCTGCTTATATGTAT 1765  
 |||||  
 QY 1795 gaacacggttaggcatcaacaagacattccactctggcaggaagattgtacgacatggct 1854  
 |||||  
 Db 1766 GAACACGGCTTAGGCATACAAAGGACATTCACCTTGGCCAGAGATTTGACGACATGGCT 1825  
 |||||  
 QY 1855 gctcaaacggtccagatgccacataccctgctcttctggcgtcatgaaactggaact 1914  
 |||||  
 Db 1826 GCTCAAACGAGTCCAGATGCCACATACCTGTCTTTTGGCGTATGAACCTGGAACCT 1885  
 |||||  
 QY 1915 acgcatttctcgggatactcctgtttttaaactcagttcacacagagatggaactggctg 1974  
 |||||  
 Db 1886 ACGCATTTGCTCGGGATATCTGTTTTTAAT---TTCAACACGAGATGGAACCTGGCTG 1942  
 |||||  
 QY 1975 aaactggacaacacattggacacactgggaacttatttggatggcctcattgttct 2034  
 |||||  
 Db 1943 AAACCTGGACAACACCATTTGGACACACTGGGACTTATTGTGATTGGCCTCATTTGTCCT 2002  
 |||||  
 QY 2035 gggctgatttggct-gcttagaataaccatgggtag 2070  
 |||||  
 Db 2003 GGCGCTGATTTGTGGCTTAGAAATCACCATGGGTAG 2039  
 |||||

RESULT 5  
 AX148141  
 LOCUS AX148141 1500 bp DNA PAT 08-JUN-2001  
 DEFINITION Sequence 9 from Patent WO0136636.  
 ACCESSION AX148141  
 VERSION AX148141.1 GI:14347063  
 KEYWORDS human.

SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1500)  
 AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.  
 TITLE Human notch ligand proteins and polynucleotides encoding the same  
 JOURNAL Patent: WO 0136636-A 9 25-MAY-2001;  
 Lexicon Genetics Incorporated (US)

## FEATURES

Location/Qualifiers  
 1..1500

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 537 a 254 c 328 g 380 t 1 others  
 ORIGIN

## Query Match

Best Local Similarity 67.8%; Score 1403.6; DB 6; Length 1500;

Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagcccttctctgttaataagatatattgataaattctctggggtcacaataaaact 60  
 |||||  
 Db 1 ATGAAGCCCTTCTCTCTGTTAATAGAGATATTGATAATTCTTTGGGGTCACAATTAATAACT 60  
 |||||  
 QY 61 atcaagcagagaagacataataaaagacaaaaggaagaaatgtcaccacacaggtatca 120  
 |||||  
 Db 61 ATCAAGCAGAGGAACATAATAAAGACAAAAGGAAGAAATGTCCACCACAGGTATCA 120  
 |||||  
 QY 121 gtgaacgaatacaacaattatttaccacatatattggaacaaagaacatcttagtaatga 180  
 |||||  
 Db 121 GTGAACGAATCAACAATATTATTATCACACATATTGGAAACAAGAACAATCTAGTAATGTA 180  
 |||||  
 QY 181 atcaataaaggaagaaatctcctggagaaagaaagaatcaacgtaaaaataagaataaaa 240  
 |||||  
 Db 181 ATCAATAAAGAGAAATCTCCTGGAGAAAAGAAAGAAATCAACGTAAAAATAAGATAAAA 240  
 |||||  
 QY 241 ggaattcaaaataaagatatcttgaagagaataaagaatcatttcaaaaagcaagcagag 300  
 |||||  
 Db 241 GGAATTCAAAATAAAGATATCTTGAAGAGAAATAGAAATCATTTACAAAAGCAAGCAGAG 300  
 |||||  
 QY 301 aaaaattttcacagatgaagagaccagctatttaagatgggcatcaagggtctctccagcag 360  
 |||||  
 Db 301 AAAAATTTTACAGATGAAGAGACCAGCTATTTAAGATGGGCATCAAGGTTCTCCAGCAG 360  
 |||||  
 QY 361 tctaaaagccaaacaaagaaagcctacctacttttgcacaaagcagctgacatg 420  
 |||||  
 Db 361 TCTAAAAGCCAAACAAAGAAAGAAAGAGCTACTACTTTTGGCCAAAGCAGCTGACATG 420  
 |||||  
 QY 421 gaaacttgaagctatggagaaaatgctgacgtcttctgatttgggaatttggcgtg 480  
 |||||  
 Db 421 GGAACTTTGAAAGCTATGGAGAAAATGGCTGACGCTTCTGCTATTGGAAATTTGGCGTG 480  
 |||||  
 QY 481 caaaataaacgcagctatccaattatatagctccttgccttaagaaagagatcatgtaaa 540  
 |||||  
 Db 481 CAAATATAACAGCAGCTATCCAAATTATATAGTCTCTGGCTAAAGAAGCATCATGTAAA 540  
 |||||  
 QY 541 gcccaaacgcattaggattttgtcttcttatggaatagggaatggaatgatcatcaagct 600  
 |||||  
 Db 541 GCCCAAAACGCATTAGGATTTTGTCTTCTTATGGAATAGGAATGGAATATGATCAAGCT 600  
 |||||  
 QY 601 aaggcaactgatattacacctttgggaagtgctggaggaacacatgatgtcccagatgatt 660  
 |||||  
 Db 601 AAGGCACGTATATATACACCTTTGGAAGTGTGGAGGAACATGATGTGCCAGATGATT 660  
 |||||  
 QY 661 ttgggttacagatatattgtcgggaatcaatgttctacagaatttggaagtgccttaagt 720  
 |||||  
 Db 661 TTGGGGTACAGATATTTGCGGGAATCAATGTCTACAGAAATGTGGAAGTTGCCCTTAAGT 720  
 |||||  
 QY 721 tattacaagaagtgagcagattatttgcgcacacatttgaacaaagtgaaggtgttcca 780  
 |||||  
 Db 721 TATTACAAGAAAGTGGCAGATTATATTGCTGACACATTTGAAAAAAGTCAAGGTGTTCCTCA 780  
 |||||  
 QY 781 gtcgaaaaagtgagactaacggaagaccctgaaatctctagttcttaacagtgagatttgc 840  
 |||||  
 Db 781 GTGGAAGAAAGTGAAGACTTAACGGAAAGACCTGAAAAATCTGAGTTCTAACAGTGAATTTG 840  
 |||||  
 QY 841 gattggacatacaataactactataaattttggcagaagaagagagatttcagatcaca 900  
 |||||  
 Db 841 GATTGGACATATACCAATACATATAAATTTTGGCAGAAAGAGGAGATGTTTCAGATACAA 900  
 |||||  
 QY 901 gtctctcttggacaattacatcaatttggcaggaaggtctagatcaggtactactacaaa 960  
 |||||

```
Db 901 GTCTCTCTGGCAATTAACATCTAATTCGGAGGAAGCTCTAGATCAGGATTAACAAA 960
Qy 961 gcattacactcttctaaagcagcaagcccgaggatgcaaatgccatgcatttata 1020
Db 961 GCATTACACTACTCTTAAGGCGCAAGAGCCGGGAGTGCAATGCCATGGCATTTATA 1020
Qy 1021 ggaagatgtatttagaggggaatgctccgtgccgcaaaaataacgcctactgcctcaag 1080
Db 1021 GAAAGATGTATTTAGAGGGAAATGCTGCCGTGCCGCAAAATAACGCCTACTGCCCTCAAG 1080
Qy 1081 tacttttccatggcagcagtaagggaatgcaatgcctccttcattgccttggtcttctt 1140
Db 1081 TACTTTTCCATGGCAGCAGTAAGGGCAATGCAATCGGCCCTCAATGGCCTTGGCTCTCT 1140
Qy 1141 tacttctatgaaagaggttcccttgatctatgcraagcacttaatactttcagaaa 1200
Db 1141 TACTTTCATGGAAAGAGGTTCCTCGTGAATATGCCRAGGCATTAATATCTTTCAGAAA 1200
Qy 1201 gctcggaagaaagggtggccgcagcagctccagttccagtttagcttcatctctggc 1260
Db 1201 GCTCGGAAAAAGGTTGGCCGACGACAGTTCAGTTAGGCTTCATCTACTCTGGC 1260
Qy 1261 tctggaatatggaagattataaaacttgccttcaaatatttttaacctggcctcagagt 1320
Db 1261 TCTGGAATATGGAAGGATTATAAACTTGCCTTCAAAATATTTTACTGGCATCTCAGAGT 1320
Qy 1321 gggcgccccctgcctatttatctgcccagatgtgcaacagggaacaggagtagta 1380
Db 1321 GGGAGCCCCCTGCCATTTATTTCTGCCAAGATGTATGCAACAGGAACAGGAGTAGTA 1380
Qy 1381 agatcatgcagaactgctgtggag 1404
Db 1381 AGATCATGCAGAACTGCTGTGGAG 1404

RESULT 6
LOCUS AX148145 1728 bp DNA PAT 08-JUN-2001
DEFINITION Sequence 13 from Patent WO0136636.
ACCESSION AX148145
VERSION AX148145.1 GI:14347065
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1728)
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.
TITLE Human notch ligand proteins and polynucleotides encoding the same
JOURNAL Patent: WO 0136636-A 13 25-MAY-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
source 1..1728
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 592 a 309 c 380 g 446 t 1 others
ORIGIN

Query Match 67.8%; Score 1403.6; DB 6; Length 1728;
Best Local Similarity 100.0%; Pred. No. 2.8e-268;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaagcccttgcctgtctgttaagagattgataaattcttggggtcacaaataaaact 60
Db 1 ATGAAGCCCTTGTCTCTTTATAGAGATATTGATTAATCTTGGGGTCACAAATPAAACT 60
Qy 61 atcaagcagagggaacataataaaagacaaaaggaagaaatgtccaccacagagtatca 120
Db 61 ATCAAAGCAGAGGAACATAATAAAAGACAAAGGAAGAAATGTCCACCACACAGGTATCA 120
Qy 121 gtgacgaaatcaacaatatttatcacacatattggacaaagaacatctagtaatgta 180
```

```
Db 121 GTGAACAAATCAACAATATTTATCACACATATTTGGAACAAAGAACATCTAGTAATGTA 180
Qy 181 atcaataaaagaaaaatctctctgagaaaaagaagaatcaacgtataaataagaataaaa 240
Db 181 ATCAATAAAAGAGAAAATCTCTCTGGAGAAAAGAAGAAATCAACGTAAAATAGAATAAAA 240
Qy 241 ggaattcaaaataaagatatcttgaagagaaaataagaatacatcttcaaaaaagacagag 300
Db 241 GGAATTCAAAATAAAGATATCTTGAAGAGAAAATAAGAATCATATTTACAAAAGCAAGCAGAG 300
Qy 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaagggttctccagcag 360
Db 301 AAAAATTTTACAGATGAAGGAGACCAGCTATTTAAGATGGGCATCAAGGTTCCTCCAGCAG 360
Qy 361 tctaaaagcccaaaacaaaaagaagaccctaccttacttttgcacaaagcagctgacatg 420
Db 361 TCTAAAGCCCAAAACAACAAAGAGAACGCCCTACTCTACTTTTTGGCCAAAGACAGCTGCATG 420
Qy 421 ggaacttgaagctatggagaaaaatggctgacgccttctgtctatttggaaaattttggcgtg 480
Db 421 GAAACTTTGAAAGCTATGGAGAAAATGGCTGACGCTTTGCTATTTGGAAAATTTTGGCGGTG 480
Qy 481 caaaataaaacagcagctatccaaattatagtgcttggctaaagaaggatcatgtaaa 540
Db 481 CAAAATATAACAGCAGCTATCCAAATATATGAGTCCCTTGGCTAAAGAAAGGATCATGTAAA 540
Qy 541 gcccaaacgccttaggattttgtcttcttatggaataggaatggaatatgatacaagt 600
Db 541 GCCCAAAACGCATTAGGATTTTGTCTTCTTATGGAANTAGGAATGGAATATGATCAAGCT 600
Qy 601 aaggcactgatatttacaccttgggaagtgctggaggaacacatgatgtcccagatgatt 660
Db 601 AAGGCATGATATATTACACCTTTTGGAAAGTCTGGAGGAAACATGATGTCCCAGATGATT 660
Qy 661 ttgggtcacagatatgttcgggaataatgtttctacagaatttggaagttgccttaagt 720
Db 661 TTGGGGTACAGATATTTCTCGGGAATCAATGTCTACAGATTTGGAAGTTGCCCTTAAGT 720
Qy 721 tattacaagaagtgcgagatttatattgctgacacatttgaaaaaagtgaaggtgttcca 780
Db 721 TATTACAGAAGAAGTGGCAGATTATATTGCTGCACATTTGAAAAAAGTGAAGGTGTTCCTCA 780
Qy 781 gtgaaaaagtgcagtaacgcgaaagacctgaaactctgagttctaaacagtgagatttg 840
Db 781 GTGGAAGAAAGTCAGACTAAACGGAAGACCTGAAAATCTGAGTCTAACAGTGAGATTTTG 840
Qy 841 gatgggcataaccataactataataatttttggcagagaagaggagatttcagatacaa 900
Db 841 GATTGGGACATATACCAATACTATATAATTTTGGCAGAAAGAGAGATGTTTCAGATACAA 900
Qy 901 gtctctctggacaattacatctaattggcaggaaggtctctagatcaggattactacaaa 960
Db 901 GTCTCTCTTGGCAATTAACATCTAATTTGGCAGAAAGGTCTAGATCAGGATTAACAAA 960
Qy 961 goattacactacttctaaagcagcaagccggggagtgcaaatgccatggcatttata 1020
Db 961 GCATTACACTACTTCTTAAAGGCAGCAAGGCCGGGAGTGCATAATGCCATGTCATTTATA 1020
Qy 1021 ggaagatgtatttagaggggaatgctgccgtgcccgaataataacgcctactgcctcaag 1080
Db 1021 GGAAGATGTATTTAGAGGGGAATGCTGCCGTGCCGCAAAATAACGCCTACTGCCCTCAAG 1080
Qy 1081 tacttttccatggcagcagtaaggcgaatgcaatgcgccttctcaggttgggtcttctt 1140
Db 1081 TACTTTTCCATGGCAGCAGTAAGGGCAATGCAATCGGCCCTCAATGGCCTTGGCTCTCT 1140
Qy 1141 tactttcatgaaaaagaggttcccttgatctatgcraagcacttaatactttcagaaa 1200
Db 1141 TACTTTTCATGAAAAAGGAGTTCCCTCTGAATATATGCCRAAGCACTTAAATACTTTCAGAAA 1200
Qy 1201 gctcggaagaaagggtggccgcagcagttccagttccagtttagcttactactctggc 1260
```







KEYWORDS human.  
SOURCE ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1257)  
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.  
TITLE Human notch ligand proteins and polynucleotides encoding the same  
JOURNAL Patent: WO 0136636-A 7 25-MAY-2001;  
Lexicon Genetics Incorporated (US)  
FEATURES Location/Qualifiers  
source 1..1257  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 464 a 206 c 270 g 316 t 1 others  
ORIGIN  
  
Query Match 60.6%; Score 1254.6; DB 6; Length 1257;  
Best Local Similarity 100.0%; Pred. No. 9.4e-239;  
Matches 1255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 atgaagccctgtctcttaatacagattatgataattcttggggtcacaaattaaaact 60  
Db 1 atgaagccctgtctctgttaatacagattatgataattcttggggtcacaaattaaaact 60  
  
Qy 61 atcaagcagaggaacaataaagaacaaagaaagaaatgtcaccacacaggtatca 120  
Db 61 atcaagcagaggaacataaataaagaacaaagaaagaaatgtcaccacacaggtatca 120  
  
Qy 121 gtgaacgaatcaacaatatttatcacatatattggaacaaagaacatctagtta 180  
Db 121 gtgaacgaatcaacaatatttatcacatatattggaacaaagaacatctagtta 180  
  
Qy 181 atcaataaagaagaatctctgagaaaaagaagaatcaacgtataaagaataaaa 240  
Db 181 atcaataaagaagaatctctgagaaaaagaagaatcaacgtataaagaataaaa 240  
  
Qy 241 ggaattcaaaataaagatatcttgaagagaataagaatcatttcacaaagcaagcag 300  
Db 241 ggaattcaaaataaagatatcttgaagagaataagaatcatttcacaaagcaagcag 300  
  
Qy 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaaggtttccagcag 360  
Db 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaaggtttccagcag 360  
  
Qy 361 tctaaagcccaaaacaaagaagaacacctactcttcttttgcacaaagcaagcagatg 420  
Db 361 tctaaagcccaaaacaaagaagaacacctactcttcttttgcacaaagcaagcagatg 420  
  
Qy 421 ggaacttgaagctatggagaaaatggctgacgcttctgctatttggaaaatttggcgtg 480  
Db 421 ggaacttgaagctatggagaaaatggctgacgcttctgctatttggaaaatttggcgtg 480  
  
Qy 481 caaataataacgcagctatccaatatatagtgctcttggcttgaagaaggatcatgtaaa 540  
Db 481 caaataataacgcagctatccaatatatagtgctcttggcttgaagaaggatcatgtaaa 540  
  
Qy 541 gcccaaacgcattagattttgtctctcttaataagaagaatggaatgaatgaatgaagct 600  
Db 541 gcccaaacgcattagattttgtctctcttaataagaagaatggaatgaatgaatgaagct 600  
  
Qy 601 aaggcactgatattacaccccttggaaagtgtgaggaacacatgatgtcccagatgatt 660  
Db 601 aaggcactgatattacaccccttggaaagtgtgaggaacacatgatgtcccagatgatt 660  
  
Qy 661 ttgggggtacagatatttgcgggaatcaatgtttctacagaattgtgaagttgccctaagt 720  
Db 661 ttgggggtacagatatttgcgggaatcaatgtttctacagaattgtgaagttgccctaagt 720  
  
Qy 721 tattacaagaagaagtgccagatttatattgtgcacacatttgaaaaaagtgaaagtggtcca 780  
Db 721 tattacaagaagaagtgccagatttatattgtgcacacatttgaaaaaagtgaaagtggtcca 780

Qy 781 gtggaaaaagtgagactaacggaagacctgaaactctgagttcttaacagtgagattttg 840  
Db 781 gtggaaaaagtgagactaacggaagacctgaaactctgagttcttaacagtgagattttg 840  
  
Qy 841 gattgggacataaccataactataataatttttggcagaagaaggagagtgctcagatacaa 900  
Db 841 gattgggacataaccataactataataatttttggcagaagaaggagagtgctcagatacaa 900  
  
Qy 901 gtctctcttgacaaattacatctaattgacgagaaagctctagatcagagattactacaaa 960  
Db 901 gtctctcttgacaaattacatctaattgacgagaaagctctagatcagagattactacaaa 960  
  
Qy 961 gattacactacttcttaagcgacgaagcgccgggagtgcaaatgccatggcatttata 1020  
Db 961 gattacactacttcttaagcgacgaagcgccgggagtgcaaatgccatggcatttata 1020  
  
Qy 1021 ggaagaatgatttagaggggaatgctgccgtgcgcgaataaaccgctactgcctcgaag 1080  
Db 1021 ggaagaatgatttagaggggaatgctgccgtgcgcgaataaaccgctactgcctcgaag 1080  
  
Qy 1081 tactttccatggcagccagtaaggcgaatgaatcgcccttcattggcttgctctct 1140  
Db 1081 tactttccatggcagccagtaaggcgaatgaatcgcccttcattggcttgctctct 1140  
  
Qy 1141 tactttcatggaaaaaggagttcccttgatattgcccraagcaacttaatactttcagaaa 1200  
Db 1141 tactttcatggaaaaaggagttcccttgatattgcccraagcaacttaatactttcagaaa 1200  
  
Qy 1201 gctgcggaagaagggtgcccgcagcacagttccagttcaggtttagcttactact 1255  
Db 1201 gctgcggaagaagggtgcccgcagcacagttccagtttccagtttagcttactact 1255

RESULT 9  
E26632 LOCUS E26632 2382 bp DNA PAT 07-FEB-2001  
DEFINITION TSA305 gene.  
ACCESSION E26632  
VERSION E26632.1 GI:13018167  
KEYWORDS JP 199215987-A/1.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 2382)  
AUTHORS Yosuke,H.K.O.O.  
TITLE TSA305 gene  
JOURNAL Patent: JP 199215987-A 1 10-AUG-1999;  
COMMENT OTSUKA PHARMACEUT CO LTD  
OS Unidentified  
PN JP 199215987-A/1  
PD 10-AUG-1999  
PF 20-APR-1998 JP 1998126803  
PR  
PI YOSUKE HARADA,KOICHI OZAKI  
PC C12N15/09,C07K14/47,C07K16/18//C07K14/435,C12N5/10,C12Q1/68,  
PC C12N15/00,  
PC C12N5/00  
CC Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..2382 /organism='Unidentified'.  
FEATURES Location/Qualifiers  
source 1..2382  
/organism='unidentified'  
/db\_xref='taxon:32644'  
BASE COUNT 689 a 501 c 614 g 578 t  
ORIGIN

Query Match 30.5%; Score 631.4; DB 6; Length 2382;  
Best Local Similarity 60.5%; Pred. No. 2.6e-115;

Matches 1056; Conservative 0; Mismatches 686; Indels 3; Gaps 1;

Qy	285	acaaagcaagcagagagaaaaattttacagatgaagagaccagcagctatttaagatgaggcat	344
Db	513	AGAAAGGCTGCTTAAGAGACGCAGATGCAGGAAGCAAAATGATGATCAAACTGGAAAT	572
Qy	345	caaggtctccagcagctcaaaagccaaacaaagaaagaaagcctaccactcttttgc	404
Db	573	GAATACTCTTAATGGAGGCAATTAAGAAAAGCCAAAAGAGAAAGCATATCGGTATCTCCA	632
Qy	405	caagcagctgacatgggaaaacttgaaagctatggagaaaaatggctgacgctttgctatt	464
Db	633	AAAGGCAGCAAGCATCAACCATACCAAAAGCCCTGGAGAGAGTGTCATATGCTCTTTTATT	692
Qy	465	tggaaaatttggcgtgcaaaaataaacaagcagcgtatccaaatatatagctccttggtctaa	524
Db	693	TGGTGATTACTTGCCACAGAAATATCAGGACGCGAGAGAGATGTTTGAGAAGCTGACATGA	752
Qy	525	agaaggatcatgtaaaagcccaaacgcattagattttgtctctctatggaataggaaat	584
Db	753	GGAAAGGCTCTCCCAAGGGACAGACTGCTTCTGGCTTTCTGTATGCCCTCGACTTGGTGT	812
Qy	585	ggaatatgatcaagctaaagcactgatattatacaccttttggaaagtctctggagaaaaaat	644
Db	813	TAATTCAAAGTCAGGCAAAAGGCTCTTGATATATATATACATTTGGAGCTCTTGCGGGCAATCT	872
Qy	645	gatgtccagatgattttgggtacagatattgttcgggaatacaatgtctctacagaattg	704
Db	873	AAATGCCACATCGTTTTGGTTTACAGATACTTGGCTGGCATCGGCTCTCCAGAGTTG	932
Qy	705	tgaagtgcctaaagtattacaagaaaagtggcagattattgtcacacatttgaaaa	764
Db	933	TGAATCTGCCCTGACTACATATCGTCTGTTGCCAATCATGTTGCTAGTGATATCTCGCT	992
Qy	765	aagtgaaggtgtccagtggaagaaagtgaagactaaagcgaagacctaaaatctgagttc	824
Db	993	AACAGGAGGCTCAGTAGTCAGAGAAATACGGCTGCCTGATGAATGGAAATCCAGGAAT	1052
Qy	825	taecagtgagattttggattgggacataaccaatactataaaatttttggcagaagaagg	884
Db	1053	GAACAGTGGAAATGCTAGAGAAGATTGTATTCAATATTACCAGTTCTTAGCTGAAAAAGG	1112
Qy	885	agatgttcagatacaagctctctcttgacaaatcacatctaatttggcaggaaaaggtctaga	944
Db	1113	TGATGTACAAGCAAGGTTGGTCTTTGGACAAGCTGCACCTGCACGGAGGGCGTGGAGTAGA	1172
Qy	945	tcaggattactacaagagcttaccactactctttaaaggcacaagagcgggagtgcaaa	1004
Db	1173	ACAGAATCATCAGAGACATTTGACTACTCTCAATTTACGACCAAAATGCTGCAATTCACA	1232
Qy	1005	tgccatggcatttataggaagaatgatatttagagggggaatgctgcgtgcgcgcaaaataa	1064
Db	1233	TGCCATGGCCCTTTTGGGAAGATGTATTTCGGAAGAGAGTGGACATTTGTACCTCAGAGTAA	1292
Qy	1065	cgctactgccttcaagactttttccatggccagccagtaaggcgaatcaatcgcccttca	1124
Db	1293	TGAGACAGCTCTCCACTACTTTAAGAAAGCTGCTGACATGGGCAACCCAGTTGGACAGAG	1352
Qy	1125	tgggcttggctcttcttactttcatggaaaaggagttcccccgaattatgcrcraagcaat	1184
Db	1353	TGGGCTTTGGAATGGCCTACTCTATGGGAGAGAGTTCAAAGTTAATATGATCTAGCCCT	1412
Qy	1185	taaaactttcagaagaagctcgggaaaaaagggtggcccgacgcacagttccagttaggcctt	1244
Db	1413	TAAGATTTTCCAGAAAGCTGCTGAACAGAGCTGGGTGGATGGGCAGCTACAGCTTGGTTC	1472
Qy	1245	catgtactactctggctctggaaatatggaaggaattataaaacttgccctcaaatatttta	1304
Db	1473	CATGCTACTATAATGGCAATTTGGAGCTCAGAGAGATTTAAACAGGCCCTTGAGTATTTTAA	1532
Qy	1305	ccctggcatctcagagtgggcagcccccctggccatttattatctgcccagaatgtatgcaac	1364
Db	1533	TTTAGCTCTTCAGGAGGCGCATATCTTGCTTTCTATTAACCTAGCTCAGATGATGCCAG	1592

Qy	1365	aggacagaggagtagtaagatacatatgcagaactgctgtggaggtttataaaagggtgctctgca	1424
Db	1593	TGGCACCGGCGTGATCGGATCATGTACACATCGACTGGGAGTTGTTTAAGAAATGTATGTGA	1652
Qy	1425	actagggcactgggctgaaataactctgacgaacttaacttgcctataaagatgctgatat	1484
Db	1653	ACGAGGCGGTGTGCTGTGAAGGCTTATGACTGCCCTATTAACAGCTATTAAGATGCGGATTA	1712
Qy	1485	agattctctcttcttgcagtatgcactgcttgcagaaatgggggatgaagttagctcaag	1544
Db	1713	CAATGCTGCAGTGATCCAGTACCTCTCCTGCTGAACAGGGCTATGAAGTGGCACAAAG	1772
Qy	1545	caattcgaatctcatttggaaatctataaaaggctaacattcttggaaaaagagaagtgtga	1604
Db	1773	CAATGCGACGCTTTATTTCTTGATCAGAGAGAAGCAAGCAATGTGTAGTGAGAAATGAACCTTA	1832
Qy	1605	tccaatggggctctcctcatgaaatcgagctgccattcaaggccaatgcatttgcctagagt	1664
Db	1833	TCCCAGAGCTTTGCTACATTTGNAACAGGGCCGCTCTCAAGGCTATACTGTGGCTAGAAAT	1892
Qy	1665	aaaaattggagattaccattactatgtgtatgggtactgaagaaagactatcaaacagcagc	1724
Db	1893	TAAGCTCGGAGACTACCATTTCTATGGGTTTGGCCACCGATGTAGATTATGAAACTGCATT	1952
Qy	1725	cacactacagcatttgcagcccaacaaatacacacaacgcgcaagcccatgttccaatctggc	1784
Db	1953	TATTCAATTACCGTCTGGCTCTTGAGCAGCAACAGTAGTCACAGCAATGTGTTTAATCTGGG	2012
Qy	1785	ttatatgtatgaacacggcttaggcatacacaaaggacattccaattggccagaagattgtga	1844
Db	2013	ATATATGCATGAGAAAGGACTGGGCATTAAACAGGATATTACCTTGGCAACGCTTTTGA	2072
Qy	1845	cgacatggctgtctcaaacagagtcagatgccacaataactgtgctcttggcgcgtcatgaa	1904
Db	2073	TGACATGGCAGCTGAAAGCCAGCCAGATGCACAAGTCTCCAGTCTTCTTAGCCCTCTGCAA	2132
Qy	1905	acttgaaaactacgcatttctgcctcgggatactcgttttttaatcagttccacaacgagatg	1964
Db	2133	ATTGGGCGTCGCTCTATTCTTCGCTACATACGG--GAAACAACATTCGAGATATGTT	2189
Qy	1965	gaactgggtgaaactggacaacacacattggaccacactgggaacttatttggatggcct	2024
Db	2190	CACCCAACTTGATATGGACAGCAGCTTTTGGGACCTCTGAGTGGGACCTTTACCTCATGACCAT	2249
Qy	2025	catgg 2029	
Db	2250	CATTG 2254	
RESULT 10			
AF052059			
LOCUS	AF052059	3473 bp	mrna
DEFINITION	Homo sapiens	SELL1L (SELL1L) mRNA, complete cds.	PRI 02-FEB-2000
ACCESSION	AF052059		
VERSION	AF052059.1	GI:6851088	
KEYWORDS			
SOURCE			human.

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 3473)
AUTHORS	Blunio, I., Cattaneo, M., Volorio, S. and Zollo, M.
TITLE	Direct Submission
JOURNAL	Submitted (03-MAR-1998) ITBA, CNR, Via Fratelli Cervi 93, Segrate MI 20090, Italy
FEATURES	Location/Qualifiers
source	1. .3473
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="14"
	/map="14q31"
	/dev_stage="adult"



Db 2283 CACCCAACTTGATATGACCAGCTTTTGGGACCTGAGTGGGACCTTTTACCTCATGACCAT 2342  
QY 2025 cattg 2029  
|||||  
Db 2343 CATTG 2347  
RESULT 11  
E26633 7885 bp DNA PAT 07-FEB-2001  
LOCUS TSA305 gene.  
DEFINITION TSA305 gene.  
ACCESSION E26633  
VERSION E26633.1 GI:13018168  
KEYWORDS JP 199215987-A/2.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 7885)  
AUTHORS Yosuke HARADA, KOICHI OZAKI  
TITLE TSA305 gene  
JOURNAL Patent: JP 199215987-A 2 10-AUG-1999;  
OTSUKA PHARMACEUT CO LTD  
COMMENT OS Unidentified  
PN JP 199215987-A/2  
PD 10-AUG-1999  
PF 20-APR-1998 JP 1998126803  
PR YOSUKE HARADA, KOICHI OZAKI  
PI C12N15/09, C07K14/47, C07K16/18, C07K14/435, C12N5/10, C12Q1/68,  
PC C12N15/00,  
CC C12N5/00  
CC Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT CDS 46..2427.  
FEATURES  
source  
1..7885  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2332 a 1466 c 1619 g 2468 t  
ORIGIN  
Query Match 30.5%; Score 631.4; DB 6; Length 7885;  
Best Local Similarity 60.5%; Pred. No. 2.5e-115;  
Matches 1056; Conservative 0; Mismatches 686; Indels 3; Gaps 1;  
QY 285 acaaaagcagcagagaaattttacagatgaagagaccagctattttaagatgggcat 344  
|||||  
Db 558 AGAAGAGCTGCTAAGAGACGGCAGATGCGAGGAGCAGAAATGATATCAACTGGAAAT 617  
QY 345 caaggttctccagcagcttaaaagcacaacaaagaaagagcctaccctcttttgc 404  
|||||  
Db 618 GAAATCCCTTAATGGAAGCAATAAGAAAGCCAAAGAGAGCATATCGGTATCTCCA 677  
QY 405 caaagcagctgacatgggaaacttgaagactatgagaaaaatggcttgcttgcatt 464  
|||||  
Db 678 AAGAGGCAAGCAAGTGAACATACCAAGCCCTGGAGAGAGTGTATATGCTTTTATT 737  
QY 465 tggaaatttggcggcaaaataataacacagcagctatcccaattatagcttcttgctaa 524  
|||||  
Db 738 TGGTGATTAATGCGCCACGAATATCCAGCGCAGGAGAGATGTTTGAGAAAGCTGACTGA 797  
QY 525 agaagatcatgttaaagcccaaaacgcataggatttttcttcttcttgaataggaaat 584  
|||||  
Db 798 GGAAGGCTCTCCCAAGGGACAGCTGCTTGGCTTCTGATGCTCTGGACTTGGGTGT 857  
QY 585 ggaatatgatcaagctaagcagcactgatataattacaccttggaaagtgcgtggagaaacat 644  
|||||  
Db 858 TAATTCAGTCAGGCAAGAGGCTCTGTATATATATATATATATATATATATATATATAT 917  
QY 645 gatgtccagatgatttgggttacagatatatttgcgggaacaatgttctacagaaattg 704  
|||||

Db 918 AATAGCCCATATGGTTTGGGTTACAGATACTGGGCTGGCATCGGGCTCTCCACAGATTG 977  
QY 705 tgaagttgcctaaagtattacaagaagagtgagcagattattgtcgacacatttgaaaa 764  
|||||  
Db 978 TGAATCTGCCCTGACTCACTATCGTCTTGTGCCAATCATGTGTGTAGTATATCTCGCT 1037  
QY 765 aagtgaaggtgttccagtgaaaaagtggagactaaacggaagacacctgaaaaatcgagttc 824  
|||||  
Db 1038 AACAGGAGGCTCAGTAGTACAGAGAAATACGGCTGCCTGATGAAGTGAAATCCAGGAAT 1097  
QY 825 taacagtgagatttggattgggacataatacctaactataaaatttttggcagaagaagg 884  
|||||  
Db 1098 GAACAGTGAATGCTAGAAAGAAATTTGAATTCATATTTACCACTTCTAGCTAGTAAAAAG 1157  
QY 885 agatgtcagatacaagtcctctctggacaattacatactaaatttggcagaagaaggtcaga 944  
|||||  
Db 1158 TGATGTACAGACACAGTTGGTCTTGGACAACTGCACCTGCACGGAGGGCGTGGAGTAGA 1217  
QY 945 tcaggattactacaaagcattacactactctttaaaggcagcaaaagccgggagtgcaaa 1004  
|||||  
Db 1218 ACAGAAATCATCAGAGAGCATTTGACTACTTCAATTTAGCAGCAAAATGCTGGCAATTCACA 1277  
QY 1005 tgcctatggaatttataggaagaatgtatttagaggggaatgctgcogtgcgcgaataaa 1064  
|||||  
Db 1278 TGCCATGGCCTTTTGGGAAAGATGTATTTCGGAAGGAAGTACATTTGTACCTCAGAGTAA 1337  
QY 1065 cgctactgcctcaagctactttccatggcagccagcaagggcaatgcgaatcgcccttca 1124  
|||||  
Db 1338 TGAGACAGCTCTCCACTACTTTAAGAAAGCTGCTGACATGGGCAACCCAGTTGGACAGAG 1397  
QY 1125 tgggcttggcttcttacttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1184  
|||||  
Db 1398 TGGGCTTGGAAATGGCCTACCTCTATGGGAGAGAGTCAAGTTAAATATATGATCTAGCCCT 1457  
QY 1185 taaaactttcagaagctgcggaaaaaagggtggccgcagcacagttccagttaggctt 1244  
|||||  
Db 1458 TAAGTATTTCCAGAAAGCTGCTGAACAAAGGCTGGGTGGATGGCATCAGTCTGGTTTC 1517  
QY 1245 catgtactactctgctctggaatgtgaagattataaaacttgccttcaaatattttta 1304  
|||||  
Db 1518 CATGTACTATATGGCATTGGAGTCNAGAGAGATATATAACAGGCCCTTCAAGTATTTTAA 1577  
QY 1305 cctggcatctcagatgggagcagccccctcctccttcttcttcttcttcttcttcttcttct 1364  
|||||  
Db 1578 TTTAGCTTCTCAGGGAGGCGCATATCTTGGCTTTCTATAACCTAGCTCAGATGATGCCAG 1637  
QY 1365 aggaacagagtagtaagactcagcaaaactcgtggagctttataaaaggttctcgtga 1424  
Db 1638 TGGCACCGCGTGATGCGATCATGTCACACTGAGTGGAGTGTGTTTAAAGAAATGATGTGA 1697  
QY 1425 actagccactgggttgagaaattcctgcagacttacttcttcttcttcttcttcttcttcttct 1484  
|||||  
Db 1698 ACGAGGCGCTTGGTCTGAAGGCTTATGACTGCTTATAACAGCTATAAAGATGGCGATT 1757  
QY 1485 agattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1544  
Db 1758 CAATGCTGCAGTGATCCAGTACCTCTCTCTGGCTGAACAGGCTATGAAGTGGCACAAG 1817  
QY 1545 caattcagcattcatttggaaatcaaaagcctaacttcttcttcttcttcttcttcttcttcttct 1604  
Db 1818 CAATGAGCCTTTATTTCTTTGATCAGAGAGAAAGCATTTGTAGGTGAGAAATTTA 1877  
QY 1605 tccaatggcgtctctctctatggaatcgagctgcccttcaagcaatgccttcttcttcttcttcttct 1664  
Db 1878 TCCAGAGGCTTTGCTATACATTTGGAACAGGCGCGCTCTCAAGGCTATATCTGTGGCTAGAT 1937  
QY 1665 aaaaattggagattaccattactatggctatgggactaagaagaagactatcaaacagcagc 1724  
Db 1934 TAAGCTCGAGAGACTACCAATTTCTATGGGTTTGGCCCGATGTAGATTATGAAACTGCATT 1997  
QY 1725 cacacactacagcattgagcccaacaaataccacaacgcgaagccatgttcaactctgc 1784  
Db 1998 TATTCTATTACCCTCTGGCTTCTTGAGCAGCAACACAGTGCACAAGCTATCTTTAATCTGGG 2057



QY 1245 catgtactactgtgctctggaatggaagattataaaacttgccttcaaatatttta 1304  
Db 1518 CATGTACTATAATGCGATTGGAGTCAAGAGAGATTATAAACAGGCTTGAAGTTTAA 1577  
QY 1305 cctggcatctagatggcgaccctcgcattttattatcttgcgccaagatgatgcaac 1364  
Db 1578 TTTAGCTTCTCAGGAGGCCATATCTTGCTTCTTATACCTAGCTAGATCGATGCCAG 1637  
QY 1365 aggaacagagtagtaaaagatcgagaaactctgtgagcttttataaagggtgctgtga 1424  
Db 1638 TGGCACCAGCGTGATGCGATCATGTCACACTGCAGTGGAGTTGTTTAAAGATGATGTGA 1697  
QY 1425 actaggccactgggtcgaaaaacttctgcagacttacttggctataaagatggatgat 1484  
Db 1698 ACGAGGCCCTTGGCTGAAAGCTTATGACTGCTTATACAGCTATATAAGATGGCGATTA 1757  
QY 1485 agattctcttctgttcagatgcactgcttgcagaaatggggtatgaagtgcctcaaa 1544  
Db 1758 CAATGCTGCAGTGATCCAGTACTCTCTCTGCTGAACAGGGCTATGAAGTGGCACAAG 1817  
QY 1545 caaticcagcattcattttggaatctaaaaaggctaacattcttgaagagagaagtga 1604  
Db 1818 CAATGCGAGCTTTATCTTGATCAGAGAGAACAGATGTAGGTGAGAACTTA 1877  
QY 1605 tocaatggcgctctctctatggaatcgagctgccattcaagcaatgcatttgcctagagt 1664  
Db 1878 TCCAGAGCTTTGCTTACATTTGGAACAGGGCGCTCTCAAGCTATACCTGTGCTAGAT 1937  
QY 1665 aaaaattggagattaccattactatgctatggctatgggactaagaagactatcaaacgacg 1724  
Db 1938 TAAGCTCGAGAGACTACCAATTTCTATGGGTGGCCACCGATGTAGATTATGAACACTGCAT 1997  
QY 1725 cacactacagcattgcagcacaacaaataccacaacgcgcagccatttcaactctgac 1784  
Db 1998 TATTATTAACCTCTGGCTTCTGACGAGCAACACAGTGCACAAAGCTATTTTAATCTGGG 2057  
QY 1785 ttatatgtatgacacgcgcttaggcattcaaaaggacattccttggcgcaagaagtga 1844  
Db 2058 ATATATGATGATAAGAGGACTGGCATTAACAGGATATTCACCTTGGCAACGTTTTTA 2117  
QY 1845 cgacatgctgtctaaacagatccagatgccacataacctgtgcttcttggcgcgtcatga 1904  
Db 2118 TGACATGGCAGCTGAAGCGAGCCACAGATGCACAAGTTCAGTCTTCCCTAGCCCTCTGCAA 2177  
QY 1905 actggaatacagcattgcttgcggatctctgttttttaacatcagttcacacacagatg 1964  
Db 2178 AYTGGCGTCTCTATTCTTCCAGTACATACGG---GAAACAACATTCGAGATATGTT 2234  
QY 1965 gaactgggtgaactggacaacacaccattgggaccacactggggacttatttggattggcct 2024  
Db 2235 CACCAACTTGATATGGACCACTTTTGGGACCTGAGTGGGACCTTACCTCATGACCAT 2294  
QY 2025 cattg 2029  
Db 2295 CATTG 2299

RESULT 13  
AB048195  
LOCUS AB048195 2441 bp mRNA ROD 02-SEP-2000  
DEFINITION Mesocricetus auratus mRNA for SEL1L, complete cds.  
ACCESSION AB048195  
VERSION AB048195.1 GI:9967439  
KEYWORDS  
SOURCE Mesocricetus auratus cell\_line:Hamster kidney BHK cDNA to mRNA.  
ORGANISM Mesocricetus auratus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Mesocricetus.  
1 (sites)  
Wada, M. and Moriyama, T.  
Hamster kidney SEL1L  
Published Only in Database (2000) In press

2 (bases 1 to 2441)  
Wada, M. and Moriyama, T.  
Direct Submission  
Submitted (31-AUG-2000) to the DDBJ/EMBL/GenBank databases. Tatsuya Moriyama, Kyoto University, Research Institute for Food Science; Gokasho, Uji, Kyoto 611-0011, Japan  
(E-mail: moriyama@soya.food.kyoto-u.ac.jp, Tel:81-774-38-3753, Fax:81-774-38-3752)  
Location/Qualifiers  
1..2441  
/organism="Mesocricetus auratus"  
/db\_xref="taxon:10036"  
/cell\_line="Hamster kidney BHK"  
13..2397  
/gene="SEL1L"  
13..2397  
/gene="SEL1L"  
/note="similar to Caenorhabditis elegans sel-1"  
/codon\_start=1  
/product="SEL1L"  
/protein\_id="BAB12403.1"  
/db\_xref="GI:9967440"  
/translation="MQHVGLTLLCAVLLSATSSDSDSNQSDISKSLPADES  
VKDSTTRVWAGQIFVDSDEPEVSEPLQEESESSKEEVSVEGIEIFVESPSSSK  
SYEAKRKRPVLTAEHTAGHGECHFPFLFDKEYDECTSDGRKRYRYLQKAAAMNH  
TDEKWFCTEEDAAKRMQEAEMIQAGMKILNGSNRKSQKREARYLQKAAAMNH  
TKALERSVALLFGDYLNTQIAAKEMFEKTEGSPKQATGALFLYVSGLGVSNOA  
KALYYTFGALGGLLIAMVLGYRWAGICVLOSCSVLTHRYLVANHVADISITGG  
SVVORILPDEVENPGMNSGMEEDLIQYFQVLAEGDVQVQVGLQHLHGGRSVEQ  
NHOAFDYENLAANGSNHAFLEKMYSEGSDIVPQSNETHALHYFKKAADMGVPG  
SGLMAYLYGRGIQVNYDLALYFQKAEQGVWDQLQGLGSMYNGIGVGRKYQKAL  
YFNLASOGGHILAFYLAQMHASGTGVNRSCHTAVLEFKNVCEERWSERLMAYNSY  
KGDYNAAVOYLLLAEGVEVAQSNAAFILDORATIVGENETVPRALLHNRASQ  
GYTVARKIDYHFFYFGTDYDYEAFIHYRLASEQHSQAAMFNPLVHHEKGLGIKO  
DIHLAKRFYDMAAEASPDQVFLALCKLGVVYFLQYIREANIRIDIFETLQDMDLLG  
PEWDLYMTIIALLGLTVIAYRQRHQHQPVPVPRPPGPWPAPQEGPEQPQPPQ"  
BASE COUNT 649 a 579 c 678 g 535 t  
ORIGIN  
Query Match 29.6%; Score 612.2; DB 10; Length 2441;  
Best Local Similarity 59.8%; Pred. No. 1.7e-111;  
Matches 1044; Conservative 0; Mismatches 698; Indels 3; Gaps 1;  
QY 285 acaaaagcaagcagagaaaaattttacacatgaagagaccagctatttaagatggcat 344  
Db 525 AGAAGATGCTGCAAAAAGACGGCAGATGCAGGAAGACAGAAATGATTTACGAGCTGGGAT 584  
QY 345 caaggttctccagcagctctaaaagccaaaacaaaagaagagcctacctcttttgc 404  
Db 585 GAAGATACCTGATGGGAGCAACAGAAAGACCAAAAGAGAGAGATATCGGTATCTTCA 644  
QY 405 caaagcagctgacatgggaaacttgaagctatggagaaaaatggctgacgcttggctatt 464  
Db 645 AAAGGCAGCAGGCATGAATCACACCAAGCCCTGGAGAGAGATATCCTACGCTCTCTATT 704  
QY 465 tggaaattttggcgtgcgaaaaatataaacacagcagctatccaaattatagctcttggctaa 524  
Db 705 TGGTGATTACCTTCACACAGAAATATCCAGCGGCTATCCAGAGAGATGTTTGAGAAATCGACCA 764  
QY 525 aagaagcatcatgaaagcccaaaacgcattaggatttttcttcttcttggaaataggaat 584  
Db 765 GGAAGGTCTCCCAAGGACAAACTGCTCTTGGCTTCTCTACGCTCTGGAGCTCGGTGT 824  
QY 585 ggaatatgatcaagctaaaggcactgatattacacacctttggaagtgcgtgggaaacat 644  
Db 825 TAATTCAAGTCAGGCAAAAGGCTCTTGTATATTATATCTTTTGGAGCCCTTGGAGCAATCT 884  
QY 645 gatgtccacagatgattttgggttacagatatttgcgggaatacaatgttctcagagaattg 704  
Db 885 GATAGCCCATATGGTTTGGGTACTCCGGTACTGGGCTGGCAGTCCGAGTCTCCAGAGTTG 944  
QY 705 tgaagtgcctcaagtattatacaagaagaagtggcagattatattgctgacacatttgaaaa 764

Db 945 TGAGTGGTACTGACCAATATTCGCTCTGTTGCCAACCAATGTTGTAGCGATATCCCT 1004  
Qy 765 aagtaagtggtccagtggaagaaagtgagactaaagaaagacctgaaatctgagttc 824  
Db 1005 GACCGAGGCTCAGTAGTCCAGAGAAATAGCGGTGCTGATGAAGTGGAAACCCAGGAT 1064  
Qy 825 taacagtgagatttggtgagacatacaataactataataatttttggcagaagaag 884  
Db 1065 GAACAGTGGGATGGAAGAGATTGATTCCAGTATTACCAAGTTCCCTGGCTGAGAAGGG 1124  
Qy 885 agatgtcaagatacaagtcctctctggacaattacatatattggcagaagaagtcaga 944  
Db 1125 TGACGTGAAGCACAGGTTGGTCTGGGCAACTGCATCTGCATGAGAGCGGCTGGAGTAGA 1184  
Qy 945 tcaggtactactacaagcattacactactctttaagagcagcaaaagccggagtgcaaa 1004  
Db 1185 ACAGAAATCACCAGAGAGCAATTTGACTACTCTCAACTTAGCAGCAAAATGTCGCAATCACA 1244  
Qy 1005 tgcgatgcaattataggaagatgtatttagaggggaatgctgcgtgcccgaataaa 1064  
Db 1245 TCGGATGGCTTCCTCGGAAAGATGATTCTGAAGGAAGCGACATCGTCCCTCAGAGTAA 1304  
Qy 1065 cgtactgcctcaagtaacttttccatggcagccagtaaggcaatgcaatcggcctca 1124  
Db 1305 TGAGACAGCACTTCACTACTTCAAGAAAGCCGCGACATGGGCAACCCCGTGGGACAGAG 1364  
Qy 1125 tgggttggtctcttactttcattggaagagggttcccttggaattatgcccraagcact 1184  
Db 1365 CGGGCTTGGAAATGGCTACTCTATGGAAGGGGAATTCAGTTAATATGACTGGCACT 1424  
Qy 1185 taactacttcagaagctgcggaaggggtggccgacgcacagtcacagttcagttaggctt 1244  
Db 1425 CAAGTATTTCCAGAAAGCTGTGACAGGCTGGTGGATGGCAGTCGACGTTGGCTC 1484  
Qy 1245 catgtactactcgtcctggaatgaatgaaggaattataaaactgcttcaaatatttta 1304  
Db 1485 TATGTACTACAAATGGCATTTGGAGTCAAGAGAGATTATAAGCAGGCTTGAAGTATTTAA 1544  
Qy 1305 cctggcactcagtgaggcagccctgcctatttatctgtgccaagatgatgcaac 1364  
Db 1545 TTTACATCTCAAGAGGCCATATCTGTGCTTTCTATACCTTGGCAGCATGCCGCCAG 1604  
Qy 1365 aggaacagagtagtaagatcatgcagaactcgtggaagctttataaagtgctgtga 1424  
Db 1605 CGGCACAGGGGTGATGAGTCTGTGCACACTCGGTGGAGTTGTTTAAAGACGTGTGA 1664  
Qy 1425 actagggcaactgggtgagaatctctgacagcttactttgcttataagagtgatgat 1484  
Db 1665 GCGAGGCGCTGGTCGGAGAGCTCATGACTGCCCTACACAGCTATAAGACGGGACTA 1724  
Qy 1485 agattctctctgttcagtatgcactgcttgcagaaatgggggtatgaagtgcctcaag 1544  
Db 1725 CAATGCTGAGTGGTCCAGTACTCTGCTGCGCCGAACAGGCTACGAAGTGGCAGAG 1784  
Qy 1545 caattcagcattcatttggaaatctaaagggctaaccattcttggaaagagaagtga 1604  
Db 1785 CAACGACGCTTCATCTTGACAGAGAGAACCACTTGTAGGTGAGATGAGACTTA 1844  
Qy 1605 tccaatggcgtctcctcatggaatcagtcgctccattcaaggaatgcatgttagagt 1664  
Db 1845 CCCCAGAGCTTTGCTGCAATTTGGAACAGGCGCCCTCTCAAGGTTATACCGTGGCTAGAAT 1904  
Qy 1665 aaaaattggagattaccattactatggctatgggactaagaagaactatcaaacagcagc 1724  
Db 1905 TAAACTTGGAGACTACCATTTCTATGGGTTGGCACTGATGTGGATTACGACACTGCTTT 1964  
Qy 1725 cacacactacagcattgpcagcaacaaatcaccaacgcgcgaagcattgttcaatctggc 1784  
Db 1965 TATTTCATTACCAGCTGGCTTCGGAGCAACACACAGTGCCCAAGCTATGTTTAACTGGG 2024  
Qy 1785 ttatatgtatgaacagcgttaggcatcaacaaagacattcacttggccagaagattga 1844

BASE COUNT 955 a 881 c 953 g 874 t

Db 2025 CTACATGATGAGAAGGGCCTAGGCATTAAACAGGATATTACCTTGCAAAACGCTTTTA 2084  
Qy 1845 cgcacatgctgctcctcaacgagtgccagatgccacataccctgtgctctttgctgcgtcatgaa 1904  
Db 2085 TGACATGGCAGCGAGGTAGCCAGATGCACAAGTCTCTGCTCTCTTGCACACTCGCAA 2144  
Qy 1905 actggaactacgcattgctccgggtatatacctctgtttttaactagttcaccaacagatg 1964  
Db 2145 APTGGTGTGCTCTATTCTTACAGTACATACGGAAGCAAAAT--ATCCGAGACATATT 2201  
Qy 1965 gaactggcgtgaaactggacaacacacattggaccacactggggactatttggatggcct 2024  
Db 2202 CACCAACTGGATATGGACAGCTCTTGGGACCGGAGTGGACCTTACCTCATGACCAT 2261  
Qy 2025 cattg 2029  
Db 2262 CATCG 2266

RESULT 14  
AF063095  
LOCUS AF063095 3663 bp mRNA ROD 15-JAN-1999  
DEFINITION Mus musculus SEL1L (Sel1l) mRNA, complete cds.  
ACCESSION AF063095  
VERSION AF063095.1 GI:4159994  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3663)  
AUTHORS Donoviel,D.B., Donoviel,M.S., Fan,E., Hadjantonakis,A. and Bernstein,A.  
TITLE Cloning and characterization of Sel-1l, a murine homolog of the C. elegans sel-1 gene  
JOURNAL Mech. Dev. 78 (1-2), 203-207 (1998)  
MEDLINE 99077704  
REFERENCE 2 (bases 1 to 3663)  
AUTHORS Donoviel,D.B., Donoviel,M.S., Fan,E., Hadjantonakis,A.-K. and Bernstein,A.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAY-1998) Samuel Lunenfeld Research Institute, Room 983, Mount Sinai Hospital, 600 University Avenue, Toronto, ON M5G-1X5, Canada  
FEATURES  
source 1. 3663  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
1. 3663  
/gene="Sel1l"  
76. 2298  
/gene="Sel1l"  
/note="similar to Caenorhabditis elegans SEL-1, negative regulator of Notch; contains putative PEST sequence"  
/codon\_start=1  
/product="SEL1L"  
/protein\_id="AAD05210.1"  
/db\_xref="GI:4159995"

/translation="MOVPRVRLSLLLCVLLGSAATSDDKTNOODSLDSKSLPTDE  
SVKDHRTTGKVVAGQIFVDSAEAEVESLQDEEDSSKTQEEISLESPNPSKTYEE  
LKRVRKPVLTAEAEEDAAKRRMOEAEMIYQAEKILNGSNRKSQKREARYLQKAA  
GMNHTKALERSYALLFGDYLTONIAAKEMFKLEEGSPKGTGLGFLVAGLGVN  
SSOKALVYTFGALGNLIAHMLGYRWAGIGVLQSCESALTHRYLVANHVASDIS  
LTGGSVWORIRLPDEVENPGMNSMLEDLQYQFLAEKGDVQAOVLGOLHLHGR  
GVEONHOFADYFNLAANGNSHMAFLGKMYSEGSDIVPOSETALHYFKAADMGN  
PVGSGCLMAYLYGRGVVNDLALKYFKQAABQGVQDQLGSMNYNGVGVKRDYK  
QALKYFNLAQSQGHILAFYLAQMHASGTGVMRSCHTAVELFKNVCERGRWSERLMTA  
YNSYKDEYNAAVQVYLLALQAEVSAQSNAAFLIDQREATVIGNEATYPRALLHNR  
AASQCYTVARIKLDGYHFYGFCTDVEDYETAFTHYLAHQHQAQMLNGLVYMHKGL  
GIKODIHLAKRFYDMAAASPDQAPVFLALCKLVGVYFLOYIREANRDLPTQDMD  
QLLGPWDLYLMTIALLGVIAVYRQHQDIPVPRPPRPAPPPQEQGPQPPQ



## ORIGIN

Query Match	29.3%;	Score 607.2;	DB 10;	Length 3663;
Best Local Similarity	60.0%;	Pred. No. 1.6e-110;		
Matches 1030;	Conservative	0;	Mismatches 683;	Indels 3; Gaps
QY	314	atgaaggaccagctatttaagatgggcatcaoaaggtttctccagcagttctaaagcccaaa	373	
DB	455	AGGAACGACAGATGATCTATCAGCGGGATGAAGATACTGAATGGAAGCAATAGGAAGA	514	
QY	374	acacaaagaagaagcctaccctacttttgcacaaagcagctgcacatggaaacttgaag	433	
DB	515	GCCAAAGAGAGAAGCATATCGGTACCTTCAGAAGCGCAGGAGCATGAATCACACCAAG	574	
QY	434	ctatggagaaatggctgcagcttggctatttgggaaatttggcgtgcacaaataaacag	493	
DB	575	CCCTGGAGAGAGTGCTCTATGCTCTCTTGTGTGGTGATTACTCACACAGAATATCCAGG	634	
QY	494	cagctaccaattatgatgctctggctataagaaggatcatgtaaagcccaaaacgcgt	553	
DB	635	CAGCCAAAGAGATGTTTGAAACCTGACTGGAAGAGGCTCTCCCAAAGACGACGACTGGTC	694	
QY	554	taggattttgtctctttagaataggaatggaataatgatacaagctaaagcactgat	613	
DB	695	TTGGCTTCTCTACGCTTCTGGGCTTGGTGTTAATTCAAGTCAGGCAAGGCTCTTGAT	754	
QY	614	attacacctttggaagtgcctggaggaacatgatgccagatgatttggggtagagat	673	
DB	755	ATTATACTTTCCGAGCTCTTGGAGGCAACCTGATAGCCCATGATATTTGGGTTTACCGCT	814	
QY	674	atttgtgggaatcaatgtttcacagaattgtgaagtggcctaaagtattattacaagaag	733	
DB	815	ACTGGGCTGGCATCGGAGTCTCCAGAGTTGTGAGTCGGCCTGACCCATTATTCGCTTG	874	
QY	734	tggcagatttatattgtctgcacatttgaaaaagtgaaggtgttccagtggaataagtga	793	
DB	875	TTGCCAATCATGTTGCTAGTGATATCTCCCTAACTGGAGGCTCTCTAGTCACAGAAATC	934	
QY	794	gactaacgaaagacctgaaatctgagtctacaagtgagatttgggattgggacatat	853	
DB	935	GGCTGCCGATGAAGTGGAAAAACCCGGGGATGAACAGTGGGATGCTGGAAGAAGACCTGA	994	
QY	854	accaatcactataatttggcagaagaagagatgttcagatacaagttctcttgagc	913	
DB	995	TTCACTATTACAGTTCCTAGCTGAGAGGGTGTGAGCTCCACAGCAGGTTGGCCTGGGAC	1054	
QY	914	aattacatctaatttggcaggaaggtctagatccaggattactacaagcattacactact	973	
DB	1055	AGCTGCATCTGCATGSGAGSGCTGTGGAGTAGAACAGAAATCACAGAGCGTTTGACTACT	1114	
QY	974	tcttaagcgacaaagccggagatgcacaaatgcattataggaagaagtgtatt	1033	
DB	1115	TCAACTTAGCAGCAAAATGCTGGCAATTACATGCTATGGCCTTCTCTGGAAAGATGTATT	1174	
QY	1034	tagaggggaagtctgccgtgcgcacaaataaagctactgccttcaagtaactttccatgg	1093	
DB	1175	CTGAAGGAAGTGACATCGTACCTCAGGTATATGAGCGGCACTTCACACTACTTTAAGAAG	1234	
QY	1094	cagccagtaagggcaatgcaatcgcccttcattgggcttgggtcttcttactttcatggaa	1153	
DB	1235	CTGCTGACATGGCAACCCCGTGGACAGAGCGGCTTGGAAATGSCCTACCTCTACGGAA	1294	
QY	1154	aaggagttccctgtaattatgcraagcacttaatactttcagaagaagctgcggaaaaag	1213	
DB	1295	GAGCGCTTCAAGTTAATTATGACTACCTGGGCCCTCAAGTATTTCCAGAAAGTCTGAGCAAG	1354	
QY	1214	ggtggcccgacgcacagttccagtttagcttcagtactactctctggctctggaaatgga	1273	
DB	1355	GCTGGGTGGACGGGACGTCGAGCTCGGCTCTATGTACTCAATGGCAATGGAGTCAAGA	1414	
QY	1274	aggattataaacttgccttcaaatatttttacctggcatctccagatgggcagccctctg	1333	

Db	1415	GAGATATTAAGCAGGCCTTGAAGATATTTAAATCTCGCTTCTCAAGGAGGCCATATCTTGG	14174
Qy	1334	ccatttattatctggccaagatgtatgcaacagggaacaggagtagtaagatcatgcaagaa	13933
Db	1475	CTTCTATACCTTCGCACAGATGCACGCCACAGCGGTCACAGGGGTGATCGCGTCTCTGTACAC	1534
Qy	1394	ctgctgtggagcctttataaagtgctgtgtaactaggccactgggctggagtaaaatctctga	1453
Db	1535	CTCAGTGGGAGTTGTTTAAAGATGTGTGTGACGAGGTCGCTGGTTCAGAGAGACTGATGA	1594
Qy	1454	cagcttaacttgcctataaaggatggtgataatagattctctcttcttcttgcagtatgcactgc	1513
Db	1595	CTGCCATACACAGCTATAAGGATGAGGACTACAACTGTCAGTGGTCCAGTACCTCTCTGC	1654
Qy	1514	tgcagaatagggttatgaagtagctcctaaagcaatcagcattcaatttggaaatcctaaa	1573
Db	1655	TGCTTGACGAGGGCTACGAGGTGGCGCAGACGACGCTTATCTCTCCAGCCAGAGAG	1714
Qy	1574	aggctaactcttgaataaagaagaagtgtatcccaatggcgtctctctctatggaaatcgag	1633
Db	1715	AAGCAACCAATGTAGTGAATGAATGAACCTTACCCAGAGCTTTACTGCAATTTGGAACAGG	1774
Qy	1634	ctgcattcoaaggcaatgcatttgcctagtagtaaaaaattggagattacocattactatgcct	1693
Db	1775	CCGCTCCCAAGGTTACACTGTGGCTAGAAATTAAGCTTTGGAGACTACCACCTTCTATGGCT	1834
Qy	1694	atgggactaaagaagactatacaaacagcagccacacactacagcaatgcaagccaacaaat	1753
Db	1835	TTGGCACTGATGTGGAAATATGAGACGCCATTTATTTCATTACCGCTTGGCTTCTGAGCAGC	1894
Qy	1754	accacaacgcgaagccatgttccaatctggcttatgtatgaacacagcgttaggcattca	1813
Db	1895	AGCACAGCGCCCAAGCTATGTTTAACTTGGGCTACATGCACGAGNAGGGCCTTAGGCATTA	1954
Qy	1814	caaagagacattcacttggccagaagattgtacgacatgctgcctcaaacagtgccagatg	1873
Db	1955	AACAGGACATTTACCTTGTCAAAACGCTTTATGACATTGGCAGCCGAAGCTAGCCAGATG	2014
Qy	1874	ccacatacctgtgcttttgcgcgtcatgaaactggaactaacgcaatttgcctcgggata	1933
Db	2015	CACAAGTACCTGTGTTCTCTCGCACTCTGCAAAATTAGGTGTCGTATTCTTACAGTACA	2074
Qy	1934	tccgtgtttttaatcagttccaaacagagatggaactggctgaaactggagacaacaccattg	1993
Db	2075	TACGGGAAGCAA---ACATCGAGATCATATTCACACACTGGATATGACCAGCTTTTGG	2131
Qy	1994	gaccacactgggactatttgtgattggcctcattg	2029
Db	2132	GACCCGAGTGGGACCTTTACCTCATGACCATCATTTG	2167
RESULT	15		
AX079694			
LOCUS	AX079694	557 bp	DNA
DEFINITION	Sequence 438 from Patent WO0107611.	PAT	22-FEB-2001
ACCESSION	AX079694		
VERSION	AX079694.1	GI:13159263	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1. (bases 1 to 557)		
JOURNAL	Baker, K.P., Goddard, A. and Wood, W.I.		
FEATURES	Human polypeptides and methods for the use thereof		
source	Patent: WO 0107611-A 438 01-FEB-2001;		
	Genentech, Inc. (US)		
	Location/Qualifiers		
	1..557		
BASE COUNT	212 a	81 c	120 g
ORIGIN	142 t	2 others	



Query Match		24.7%;	Score 510.4;	DB 6;	Length 557;
Best Local Similarity		97.7%;	Pred. No. 2.7e-91;		
Matches	549;	Conservative	0;	Mismatches	8;
				Indels	5;
				Gaps	3;
Qy	206	agaaaaaagaagaatcaacgtaaaaaataagaataaaaggaattcaaaataaagatatcttga	265		
Db	1	AGAAAAAGAAGAAATCAACGTAA--ATAAGATAAAGGATTCAAAATAAAGATNTCTTGA	57		
Qy	266	agagaaataaagaatcatttacaagaagcagagagagaaaaattttacagatgaagagacc	325		
Db	58	AGAGAAATAAGAAATCATTTACAAAAGC-AGCAGAGAAAAATTTTACATGAAGGAG-CC	115		
Qy	326	agctatttaagatggcgatcaaggttctccagcagctctaaaagccaaaacaaaaagaag	385		
Db	116	AGCTATTTAAGATGGCATCAAGGTTCCTCCAGCAGCTCTAAAAGCCAAAAACAAAAAGAAG	175		
Qy	386	agcctactactttttccaaaagcagctgacatgggaacttgaaaactatggagaaaa	445		
Db	176	AAGCCTACTACTTTTGGCCAAAGCAGCTGACATGGGAAACTTGAAGCTATGGAGAAA	235		
Qy	446	ggctgacgctttgctatttgaaattttgcgtgcacaaataataacacagcctatccaat	505		
Db	236	TGGCTGACGCTTTGCTATTGGAAATTTTGGGTGCAAAATATAACAGCAGCTATCCAAT	295		
Qy	506	tatatgctccttggctaaagaagatcatgtaagcccaaacgcattagatttttgt	565		
Db	296	TATATGAGTCCTTGGCTAAAGAAGGATCATGTAAAGCCCAAAACGCATTAGGATTTTGT	355		
Qy	566	cttcttatggaataggaatgaatatgatcaagctaaggcactgatattacaccttg	625		
Db	356	CTTCTTATGGAATAGGAATGGAATATGATCAAGCTAAGGCACTGATATATTACGCTTTG	415		
Qy	626	gaagtgcctggagaaacatgatgtcccagatgatttgggtcacagatatttgcggaa	685		
Db	416	GAAGTGCTGGAGGAAACATGATGTCCCAAGATGATTTGGGGTACAGATATTGTGCGGAA	475		
Qy	686	tcaatgtctacagaatttgaaagttgccctaagttattacaagaaagtgccagattata	745		
Db	476	TCAATGTCTACAGAAATTGTGAAGTTGCCCTAAGTTATTACAAGAAAGTGGCAGATTATA	535		
Qy	746	ttgctgacacatttgaaaaaag	767		
Db	536	TGCTGACACATTTGAAAAAAG	557		

Search completed: April 8, 2002, 21:31:00  
Job time: 3347 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2002, 20:42:09 ; Search time 121.25 Seconds  
(without alignments)  
14636.398 Million cell updates/sec

Title: US-09-714-882-1

Perfect score: 2070

Sequence: 1 atgaagccctgtctctgtt.....ttagaatcaccatgggttag 2070

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2069.6	100.0	2070	22 AAD06374	Novel human protei
2	2053.6	99.2	2073	22 AAD06375	Novel human protei
3	1650.2	79.7	1773	22 AAD06376	Novel human protei
4	1403.6	67.8	1500	22 AAD06378	Novel human protei
5	1403.6	67.8	1728	22 AAD06380	Novel human protei
6	1403.6	67.8	1731	22 AAD06379	Novel human protei
7	1254.6	60.6	1257	22 AAD06377	Novel human protei
8	633	30.6	7885	20 AAX76578	Human pancreas-spe
9	631.4	30.5	2382	20 AAX76577	Human pancreas-spe
10	631.4	30.5	8028	21 AAC75460	Human ORFX ORF1015
11	596.4	28.8	3666	20 AAX77013	Mouse Sel-1L splic

12	594.8	28.7	3667	20 AAX77012	Full length Mouse
13	510.4	24.7	557	22 AAF93617	cDNA encoding SRT
14	444	21.4	2144	22 AAF93859	Human cDNA encodin
15	437.6	21.1	2109	20 AAX77014	Human Sel-1L codin
16	316.4	15.3	1896	20 AAX77011	Partial Sel-1L cod
17	164	7.9	545	22 AAI16647	Probe #6580 for ge
18	164	7.9	545	22 AAI19859	Probe #545 used t
19	163	7.9	163	22 AAI25843	Probe #15776 for g
20	163	7.9	163	22 AAI52940	Probe #21626 used
21	149	7.2	936	22 AAF58252	Oligonucleotide D1
22	149	7.2	936	22 AAF58252	Oligonucleotide D1
23	149	7.2	936	22 AAF58254	Oligonucleotide D1
24	149	7.2	936	22 AAF58254	Oligonucleotide D1
25	149	7.2	936	22 AAF58257	Oligonucleotide D1
26	149	7.2	936	22 AAF58257	Oligonucleotide D1
27	149	7.2	936	22 AAF58259	Oligonucleotide D2
28	149	7.2	936	22 AAF58259	Oligonucleotide D2
29	149	7.2	936	22 AAF58262	Oligonucleotide D2
30	149	7.2	936	22 AAF58262	Oligonucleotide D2
31	149	7.2	938	22 AAF58255	Oligonucleotide D1
32	149	7.2	938	22 AAF58255	Oligonucleotide D1
33	146.4	7.1	1124	20 AAX61283	Human signal pepti
34	118.4	5.7	2225	21 AAZ52467	HTRM clone 1880692
35	117	5.7	1672	22 AAF92066	Human PRO1063 cDNA
36	117	5.7	1701	21 AAZ64982	Membrane-bound pro
37	117	5.7	1701	22 AAF44128	Human PRO4063 (UNQ
38	90.2	4.4	571	18 AAT85270	Expressed sequence
39	85.4	4.1	538	22 AAI17986	Probe #7919 for ge
40	85.4	4.1	538	22 AAI42976	Probe #11662 used
41	83.2	4.0	34001	22 AAF28531	Genomic fragment #
42	80.2	3.9	6644	20 AAX33181	Base sequence of t
43	80.2	3.9	7372	20 AAX33182	Base sequence of t
44	80.2	3.9	7797	20 AAX33180	Cowpox virus bsr f
45	80.2	3.9	7996	20 AAX33184	Base sequence of t

#### ALIGNMENTS

RESULT 1	
AAD06374	
ID	AAD06374 standard; cDNA; 2070 BP.
XX	
AC	AAD06374;
XX	
DT	10-AUG-2001 (first entry)
XX	
DE	Novel human protein (NHP) cDNA #1, sharing similarity with Notch ligands.
XX	
KW	Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
KW	novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
KW	polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
KW	pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
KW	Parkinson's disease; stroke; vascular dementia; fat metabolism;
KW	cholesterol metabolism; coronary artery disease; gene therapy;
KW	cerebroprotective; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..2070
FT	/tag= a
FT	/product= "Novel human protein #1 (NHP)"
FT	1177
FT	/tag= b
FT	/note= "This ambiguity represents polymorphic site"
FT	1177..1179
FT	/tag= c
FT	note= "Encodes Glu"
XX	
XX	WO200136636-A2.
PN	
XX	
PD	25-MAY-2001.

XX 16-NOV-2000; 2000WO-US31373.  
XX  
XX  
PR 17-NOV-1999; 99US-0165959.  
XX  
XX (LEXI-) LEXICON GENETICS INC.  
PA  
XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
PI  
XX WPI: 2001-355635/37.  
XX  
DR P-PSDB; AAE02430.  
XX  
PT Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX  
PS Claim 2; Page 26; 39pp; English.  
XX  
CC The present sequence is a cDNA encoding novel human protein (NHP) which  
CC share structural similarity with animal Notch ligands, particularly  
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treat  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 2070 BP; 701 A; 379 C; 452 G; 537 T; 1 other;

Query Match 100.0%; Score 2069.6; DB 22; Length 2070;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2070; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagccctgtctctgttaatagagatattgataattcttgggtcacaattaaaact 60  
DB 1 atgaagccctgtctctgttaatagagatattgataattcttgggtcacaattaaaact 60  
QY 61 atcaagcagaggaacataataaaagacaaaaggaagaaatgtcaccacaggtatca 120  
DB 61 atcaagcagaggaacataataaaagacaaaaggaagaaatgtcaccacaggtatca 120  
QY 121 gtgaacgaatacaacaattatttatcacacatatgtggaacaagaacatctagtaatga 180  
DB 121 gtgaacgaatacaacaattatttatcacacatatgtggaacaagaacatctagtaatga 180  
QY 181 atcaataaagagaaatctcctggagaaagaaagaaatcaacgtataaataagaaataaa 240  
DB 181 atcaataaagagaaatctcctggagaaagaaagaaatcaacgtataaataagaaataaa 240  
QY 241 ggaattcaataaagatatcttgaagagaaataagaatcatttacaagaagcaagcagag 300  
DB 241 ggaattcaataaagatatcttgaagagaaataagaatcatttacaagaagcaagcagag 300  
QY 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaaggcttccagcag 360  
DB 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaaggcttccagcag 360  
QY 361 tctaaagccaaacaaagaaagcctacctacttttggcacaagcagctgacatg 420  
DB 361 tctaaagccaaacaaagaaagcctacctacttttggcacaagcagctgacatg 420

Db 1501 cagtgacgctgcgagaatgggtatgagtagtccaagcaatcagcattcatt 1560  
Qy 1561 ttggaatcaaaagctaaacattcttgaaaaagagaagatgtatccaatggcgtcttc 1620  
Db 1561 ttggaatcaaaagctaaacattcttgaaaaagagaagatgtatccaatggcgtcttc 1620  
Qy 1621 ctatggaatcagctgccattcaaggcaatgcatttgcctagtagtaaaatggagattac 1680  
Db 1621 ctatggaatcagctgccattcaaggcaatgcatttgcctagtagtaaaatggagattac 1680  
Qy 1681 cattactatgctatggactaaagaagactatcaaacagcagcgcacacactacagcatt 1740  
Db 1681 cattactatgctatggactaaagaagactatcaaacagcagcgcacacactacagcatt 1740  
Qy 1741 gcagccaacaataaccacacgcgcgaagccatgttccaatctggtcttatgtatgaacac 1800  
Db 1741 gcagccaacaataaccacacgcgcgaagccatgttccaatctggtcttatgtatgaacac 1800  
Qy 1801 ggcttaggcatacaaaagacattcaacttggccagaagattgtacgacatggctgctcaa 1860  
Db 1801 ggcttaggcatacaaaagacattcaacttggccagaagattgtacgacatggctgctcaa 1860  
Qy 1861 acgagtcagatgccacatacctgtcttggcgtcgaactgaaactggaactacgcat 1920  
Db 1861 acgagtcagatgccacatacctgtcttggcgtcgaactgaaactggaactacgcat 1920  
Qy 1921 ttgctccgggatctctgttttaatacagttccacacgagatgaaactgctgaaactg 1980  
Db 1921 ttgctccgggatctctgttttaatacagttccacacgagatgaaactgctgaaactg 1980  
Qy 1981 gacacacattggaccacactgggaactatttggattggcctcattgttctctgggctg 2040  
Db 1981 gacacacattggaccacactgggaactatttggattggcctcattgttctctgggctg 2040  
Qy 2041 atttggcttagaataccacatgggtag 2070  
Db 2041 atttggcttagaataccacatgggtag 2070

RESULT 2

AD06375  
ID AD06375 standard; cDNA; 2067 BP.

XX AC AAD06375;

XX DT 10-AUG-2001 (first entry)

XX DE Novel human protein (NHP) cDNA #2, sharing similarity with Notch ligands.

XX KW Human; nootropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
KW cholesterol metabolism; coronary artery disease; gene therapy;  
KW cerebroprotective; ss.

OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX CDS 1..2067

FT FT /\*tag= a  
FT FT /product= "Novel human protein #2 (NHP)"

FT FT 1177

FT FT /\*tag= b

FT FT /note= "This ambiguity represents polymorphic site"

FT FT 1177..1179

FT FT /\*tag= c

FT FT note= "Encodes Glu"

XX PN W0200136636-A2.

XX

PD 25-MAY-2001.

XX PF 16-NOV-2000; 2000WO-US31373.

XX PR 17-NOV-1999; 99US-0165959.

XX PA (LEXI-) LEXICON GENETICS INC.

XX PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

XX DR WPI; 2001-355635/37.

XX P-PSDB; AAE02431.

XX Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer

XX PS Claim 3; Page 28-29; 39pp; English.

XX The present sequence is a cDNA encoding novel human protein (NHP) which  
CC share structural similarity with animal Notch ligands, particularly  
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.

XX SQ Sequence 2067 BP; 700 A; 378 C; 451 G; 537 T; 1 other;

Query Match 99.2%; Score 2053.6; DB 22; Length 2067;  
Best Local Similarity 99.9%; Pred. NO. 0;  
Matches 2067; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 1 atgaagccctgtctgttaataagatattgataattcttgggtcacataaaact 60

Db 1 atgaagccctgtctgttaataagatattgataattcttgggtcacataaaact 60

Qy 61 atcaagcagaggaacataataaagacaaaaggaagaaatgtccacacaggtatca 120

Db 61 atcaagcagaggaacataataaagacaaaaggaagaaatgtccacacaggtatca 120

Qy 121 gtgaacgaatcaacaatatttatcacacatattggaacaaagaacatctagtatga 180

Db 121 gtgaacgaatcaacaatatttatcacacatattggaacaaagaacatctagtatga 180

Qy 181 atcaataaagagaataatctcctggagaaaagaaagaatacaactgaaataaataa 240

Db 181 atcaataaagagaataatctcctggagaaaagaaagaatacaactgaaataaataa 240

Qy 241 ggaattcaaaataaagatatcttgaagagaataaagaatcatttcaaaaagcagag 300

Db 241 ggaattcaaaataaagatatcttgaagagaataaagaatcatttcaaaaagcagag 300

Qy 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaagggttctccagcag 360

Db 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaagggttctccagcag 360

Qy 361 tctaaaagccaaaacaaaagaaagaaagcctacctcttttggccaaagcagctgacatg 420

Db 361 tctaaaagccaaaacaaaagaaagaaagcctacctcttttggccaaagcagctgacatg 420

```
QY 421 ggaaacttgaaagctatggagaaaaatggctgacgcgttttgcatttttggaaatttggcgtg 480
DB 421 ggaaacttgaaagctatggagaaaaatggctgacgcgttttgcatttttggaaatttggcgtg 480
QY 481 caaaatataacagcagctacccaattatatatgctgcttggctgcaaaagagatcatgtaaa 540
DB 481 caaaatataacagcagctacccaattatatatgctgcttggctgcaaaagagatcatgtaaa 540
QY 541 gccaaaaacgcattaggattttgtctctcttatgaaataggaaatggaatatatcaagct 600
DB 541 gccaaaaacgcattaggattttgtctctcttatgaaataggaaatggaatatatcaagct 600
QY 601 aaggcactgatattacacaccttggagctgctggagaaacatgatgtcccagatgatt 660
DB 601 aaggcactgatattacacaccttggagctgctggagaaacatgatgtcccagatgatt 660
QY 661 ttggggtacagatatttgcgggaatcaatgttctacagaatttgaaagttgcccctaagt 720
DB 661 ttggggtacagatatttgcgggaatcaatgttctacagaatttgaaagttgcccctaagt 720
QY 721 tattacaagaagtgagcagattatatgtctgacacatttgaaaaagtgaaagtggttcca 780
DB 721 tattacaagaagtgagcagattatatgtctgacacatttgaaaaagtgaaagtggttcca 780
QY 781 gtggaaaaagtgagactaacggaaagacctgaaaatctgagttctaacagtgagattttg 840
DB 781 gtggaaaaagtgagactaacggaaagacctgaaaatctgagttctaacagtgagattttg 840
QY 841 gattgggaacataaccaatactataataattttggcagaagaaggagatgttcagatacaa 900
DB 841 gattgggaacataaccaatactataataattttggcagaagaaggagatgttcagatacaa 900
QY 901 gtctctcttgacaattacataatgtgcagaaagctcagatcagagattactacaaa 960
DB 901 gtctctcttgacaattacataatgtgcagaaagctcagatcagagattactacaaa 960
QY 961 gcattacactactctttaaaggcagcaaaaggcgggagtgcaaatgccatggcatttata 1020
DB 961 gcattacactactctttaaaggcagcaaaaggcgggagtgcaaatgccatggcatttata 1020
QY 1021 ggaagaatgtattagagggggaatgctgcctgctgcgcgcaaaaaaaacgcactactccctcaag 1080
DB 1021 ggaagaatgtattagagggggaatgctgcctgctgcgcgcaaaaaaaacgcactactccctcaag 1080
QY 1081 tacttttccatggcagccagtaaggcgaatgcaatcgcccttcacggccttgcgttcttct 1140
DB 1081 tacttttccatggcagccagtaaggcgaatgcaatcgcccttcacggccttgcgttcttct 1140
QY 1141 tactttcatggaagagggtcccttgaaatgacraagcacttaaaacttttcagaaa 1200
DB 1141 tactttcatggaagagggttcccttgaaatgacraagcacttaaaacttttcagaaa 1200
QY 1201 gctgcggaagaaagggtggccgcagcaacagttccagttagggttcattgtactactctggc 1260
DB 1201 gctgcggaagaaagggtggccgcagcaacagttccagttagggttcattgtactactctggc 1260
QY 1261 tctggaatatgaaagattataaacttgcctcaaatatttttacctggcactcagagt 1320
DB 1261 tctggaatatgaaagattataaacttgcctcaaatatttttacctggcactcagagt 1320
QY 1321 gggcgccctccctgcatttattcttgcgcaagatgtatgcaacaggaaacaggagtagta 1380
DB 1321 gggcgccctccctgcatttattcttgcgcaagatgtatgcaacaggaaacaggagtagta 1380
QY 1381 agatcatgcagaactgctgtgagagtttataaagggtctgtgaaactaggccactgggct 1440
DB 1381 agatcatgcagaactgctgtgagagtttataaagggtctgtgaaactaggccactgggct 1440
QY 1441 gagaaattcctgacagctacttgcctataaaggatgatatagattcttctcttatt 1500
DB 1441 gagaaattcctgacagctacttgcctataaaggatgatatagattcttctcttatt 1500
```

```
QY 1501 cagtatgcactgcttgcagaaaaatggggtatgaagtagctcacaagcaattcagcattcatt 1560
DB 1501 cagtatgcactgcttgcagaaaaatggggtatgaagtagctcacaagcaattcagcattcatt 1560
QY 1561 ttggaatctaaaaaggctaaacattcttggaaaaagagaatgtatccaatggcgtctctc 1620
DB 1561 ttggaatctaaaaaggctaaacattcttggaaaaagagaatgtatccaatggcgtctctc 1620
QY 1621 ctatggaatcgagctgcccattcaaggcaatgcatttctagagtagtaaaatttgagattac 1680
DB 1621 ctatggaatcgagctgcccattcaaggcaatgcatttctagagtagtaaaatttgagattac 1680
QY 1681 cattactatggctatgggactaagaagaactatcaaacagcagccacacactacagcatt 1740
DB 1681 cattactatggctatgggactaagaagaactatcaaacagcagccacacactacagcatt 1740
QY 1741 gcagccaaacaaataccacaacgcgcaagccatgttcaatcttggttatgtatgaacac 1800
DB 1741 gcagccaaacaaataccacaacgcgcaagccatgttcaatcttggttatgtatgaacac 1800
QY 1801 ggcttaggcatacaaaaggacattcacttggccagaagattgtacgacatggctgctcaa 1860
DB 1801 ggcttaggcatacaaaaggacattcacttggccagaagattgtacgacatggctgctcaa 1860
QY 1861 acgagtcagatgccacatacctgtcttgcgtcatgaaactggaactacgcgt 1920
DB 1861 acgagtcagatgccacatacctgtcttgcgtcatgaaactggaactacgcgt 1920
QY 1921 ttgctccgggatactcgttttttaacagttcacaacagagatggaactggctgaaactg 1980
DB 1921 ttgctccgggatactcgttttttaacagttcacaacagagatggaactggctgaaactg 1980
QY 1981 gacaacacattggccacacactgggacttattgttgatggcctcattgttctctggctg 2040
DB 1981 gacaacacattggccacacactgggacttattgttgatggcctcattgttctctggctg 2040
QY 2041 attttgttcttagaaaatcaccatgggtag 2070
DB 2038 attttgttcttagaaaatcaccatgggtag 2067

RESULT 3
AAD06376
ID AAD06376 standard; cDNA; 1773 BP.
XX
AC AAD06376;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human
XX Novel human protein (NHP) cDNA #3, sharing similarity with Notch ligands.
KW Human; neurotrophic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
KW cholesterol metabolism; coronary artery disease; gene therapy;
KW cerebroprotective; ss.
XX
OS Homo sapiens.
XX
FH Key
CDS 1..1773
FT /*tag= a
FT /product= "Novel human protein #3 (NHP)"
FT misc_feature
FT 1177
FT /*tag= b
FT /note= "This ambiguity represents polymorphic site"
FT 1177..1179
FT /*tag= c
FT note= "Encodes Glu"
XX
PN WO200136636-A2.
```

XX 25-MAY-2001.  
XX  
XX  
PF 16-NOV-2000; 2000WO-US31373.  
XX  
PR 17-NOV-1999; 99US-0165959.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
XX WP: 2001-355635/37.  
DR P-PSDB; AAE02432.  
XX  
XX Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX  
XX Claim 4; Page 30-31; 39pp; English.  
XX  
XX The present sequence is a cDNA encoding novel human protein (NHP) which  
CC share structural similarity with animal Notch ligands, particularly  
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
XX Sequence 1773 BP; 607 A; 313 C; 388 G; 464 T; 1 other;

Query Match 79.7%; Score 1650.2; DB 22; Length 1773;  
Best Local Similarity 99.8%; Pred. NO. 0;  
Matches 1653; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 atgaagccctgtctctgttaatagagatatgtgataattcttgggggtcacaattaaaaact 60  
DB 1 atgaagccctgtctctgttaatagagatatgtgataattcttgggggtcacaattaaaaact 60  
QY 61 atcaagcagaggaacataataaagacacaaaggaagaaatgtccacacaggtatca 120  
DB 61 atcaagcagaggaacataataaagacacaaaggaagaaatgtccacacaggtatca 120  
QY 121 gtgaacgaatcaacaatattttatcacatatgtggaacaaagaaacatctagtatgta 180  
DB 121 gtgaacgaatcaacaatattttatcacatatgtggaacaaagaaacatctagtatgta 180  
QY 181 atcaataaagaaataatctctgagaaaaaagaaagaaatcaacgtaaaaataagaataaaa 240  
DB 181 atcaataaagaaataatctctgagaaaaaagaaagaaatcaacgtaaaaataagaataaaa 240  
QY 241 ggaattcaaaataaagatatctgtgaagagaaataagaatcatttcaaaaagcaagcagag 300  
DB 241 ggaattcaaaataaagatatctgtgaagagaaataagaatcatttcaaaaagcaagcagag 300  
QY 301 aaaaattttacagatgaaggagaccagctatttaagatggggcatcaaggttctccagcag 360  
DB 301 aaaaattttacagatgaaggagaccagctatttaagatggggcatcaaggttctccagcag 360  
QY 361 tctaaagccaaacaaagaaagaaagcctactcttttttggccaaagcagctgacatg 420  
DB 361 tctaaagccaaacaaagaaagaaagcctactcttttttggccaaagcagctgacatg 420

DB 361 tctaaagccaaacaaagaaagaaagcctactcttttttggccaaagcagctgacatg 420  
QY 421 ggaacttgaaagctatgagaaaaatgctgacgctttgctatttggaaaattttggcg 480  
DB 421 ggaacttgaaagctatgagaaaaatgctgacgctttgctatttggaaaattttggcg 480  
QY 481 caaataataacagcagctatccaatttatgatgtccttggcttaagaaggatcatgtataa 540  
DB 481 caaataataacagcagctatccaatttatgatgtccttggcttaagaaggatcatgtataa 540  
QY 541 gcccaaaacgcataggatttttcttcttatgaataggaatgaatataatcaagct 600  
DB 541 gcccaaaacgcataggatttttcttcttatgaataggaatgaatataatcaagct 600  
QY 601 aaggcactgatattacaccttttgaaagctgaggaacacatgatgtcccagatgatt 660  
DB 601 aaggcactgatattacaccttttgaaagctgaggaacacatgatgtcccagatgatt 660  
QY 661 ttggggtacagatatgttcggggaatcaatgttctacagaattgtgaagttgcccataag 720  
DB 661 ttggggtacagatatgttcggggaatcaatgttctacagaattgtgaagttgcccataag 720  
QY 721 tattacaagaagtgccagatttatattctgacacatttgaataaagtgaaagtggtcca 780  
DB 721 tattacaagaagtgccagatttatattctgacacatttgaataaagtgaaagtggtcca 780  
QY 781 gtggaagaagtgagactaacggaagacctgaaatcttgatcttaacagtgagatttg 840  
DB 781 gtggaagaagtgagactaacggaagacctgaaatcttgatcttaacagtgagatttg 840  
QY 841 gattggacataacataactataaatttttggcagaagaaggagatgttcagatacaa 900  
DB 841 gattggacataacataactataaatttttggcagaagaaggagatgttcagatacaa 900  
QY 901 gtctcttggacaattacatcaattgtgcagaaaggtctagatcagattactacaaa 960  
DB 901 gtctcttggacaattacatcaattgtgcagaaaggtctagatcagattactacaaa 960  
QY 961 gcattacactactctttaaaggcagaaagcggagtgcaaatgccatgcatattata 1020  
DB 961 gcattacactactctttaaaggcagaaagcggagtgcaaatgccatgcatattata 1020  
QY 1021 ggaagatgatttagaggggaatgctgcgtgcgcgcaaaaataacgctactgcttcaag 1080  
DB 1021 ggaagatgatttagaggggaatgctgcgtgcgcgcaaaaataacgctactgcttcaag 1080  
QY 1081 tactttccatggcagcagtaagggcaatgcaatcgcttcaatggcttggcttctct 1140  
DB 1081 tactttccatggcagcagtaagggcaatgcaatcgcttcaatggcttggcttctct 1140  
QY 1141 tactttcatggaaaaggaggttccctgaattatgcraagcacttaatactttcagaaa 1200  
DB 1141 tactttcatggaaaaggaggttccctgaattatgcraagcacttaatactttcagaaa 1200  
QY 1201 gtcggaagaaaggtggcccgacgcacagttccagtttaggttctatgtactactggc 1260  
DB 1201 gtcggaagaaaggtggcccgacgcacagttccagtttaggttctatgtactactggc 1260  
QY 1261 tctggaatatgaagattataaaactgcttcaaatatttttacctggcactctcaggt 1320  
DB 1261 tctggaatatgaagattataaaactgcttcaaatatttttacctggcactctcaggt 1320  
QY 1321 gggcagccctcgcaatttatctggcagaagtatgcaacaggaacagagtagta 1380  
DB 1321 gggcagccctcgcaatttatctggcagaagtatgcaacaggaacagagtagta 1380  
QY 1381 agatcatgcagaactgctgtgaggtttataaaggtgtctgtgaactagggcactgggct 1440  
DB 1381 agatcatgcagaactgctgtgaggtttataaaggtgtctgtgaactagggcactgggct 1440  
QY 1441 ggaataattcctgacagcttactttgctataaggaatgggtgatatagattcttctctgt 1500  
DB 1441 ggaataattcctgacagcttactttgctataaggaatgggtgatatagattcttctctgt 1500

QY 1501 cagtagcactgctgcagaaatgggtatgaagtacgtccaaagcaattcagcattcatt 1560  
 |||||  
 Db 1501 cagtagcactgctgcagaaatgggtatgaagtacgtccaaagcaattcagcattcatt 1560  
 |||||  
 QY 1561 ttggaatctaaagcgaattcttgaagaaagagaagatgtatccaatgctgtctc 1620  
 |||||  
 Db 1561 ttggaatctaaagcgaattcttgaagaaagagaagatgtatccaatgctgtctc 1620  
 |||||  
 QY 1621 ctatggaatcgagctgcattcaagcgaatgctttg 1657  
 |||||  
 Db 1621 ctatggaatcgagctgcattcaagcgaatgctttg 1657  
 |||||

RESULT 4  
 AAD06378  
 ID AAD06378 standard; cDNA; 1500 BP.  
 XX  
 AC AAD06378;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE Novel human protein (NHP) cDNA #5, sharing similarity with Notch ligands.  
 XX  
 KW Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
 KW cholesterol metabolism; coronary artery disease; gene therapy;  
 KW cerebroprotective; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..1500  
 FT /\*tag= a  
 FT /product= "Novel human protein #5 (NHP)"  
 FT 1177  
 FT misc\_feature  
 FT /\*tag= b  
 FT /note= "This ambiguity represents polymorphic site"  
 FT 1177..1179  
 FT /\*tag= c  
 FT /note= "Encodes Glu"  
 FT  
 XX WO200136636-A2.  
 XX  
 PD 25-MAY-2001.  
 XX  
 PF 16-NOV-2000; 2000WO-US31373.  
 XX  
 PR 17-NOV-1999; 99US-0165959.  
 XX  
 PA (LEXI-) LEXICON GENETICS INC.  
 XX  
 PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
 XX  
 DR WPI; 2001-355635/37.  
 DR P-PSDB; AAE02434.  
 XX  
 PT Novel isolated human polynucleotides encoding polypeptides sharing  
 PT sequence similarity with mammalian SEL-1 proteins which are negative  
 PT regulators of Notch family receptors, useful for treating diabetes,  
 PT cancer  
 XX  
 PS Claim 6; Page 33-34; 39pp; English.  
 XX  
 CC The present sequence is a cDNA encoding novel human protein (NHP) which  
 CC share structural similarity with animal Notch ligands, particularly  
 CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
 CC Notch receptors and their associated signalling pathways have been  
 CC associated with development, apoptosis, neuron growth and maintenance.  
 CC Labeled NHP probes can be used to screen a human genomic library which

CC is helpful for identifying polymorphisms, determining the genomic  
 CC structure of a given locus/allele and designing diagnostic tests. The  
 CC NHP is also useful in screening techniques for drugs which treat  
 CC symptomatic or phenotypic manifestations of perturbing the normal  
 CC function of NHP in the body. Nucleotide constructs encoding functional  
 CC NHPs, antisense, antisense molecules can be used in gene therapy  
 CC approaches for modulating gene expression such as for preventing or  
 CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
 CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
 CC such as Parkinson's disease, stroke, vascular dementia and conditions  
 CC requiring modulation of fat and cholesterol metabolism such as coronary  
 CC artery disease.  
 XX

SQ Sequence 1500 BP; 537 A; 254 C; 328 G; 380 T; 1 other;

Query Match 67.8%; Score 1403.6; DB 22; Length 1500;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagcccttctctgtaagagatattgataattcttgggtgcacattaaaact 60  
 |||||  
 Db 1 atgaagcccttctctgtaagagatattgataattcttgggtgcacattaaaact 60  
 |||||  
 QY 61 atcaagcagaggaacataataaaagacaaaagaaagaaatgtcaccacacaggtatca 120  
 |||||  
 Db 61 atcaagcagaggaacataataaaagacaaaagaaagaaatgtcaccacacaggtatca 120  
 |||||  
 QY 121 gtgacgaaatcaacaataattttacacacatttggaacaaagaaacattctgtaagtca 180  
 |||||  
 Db 121 gtgacgaaatcaacaataattttacacacatttggaacaaagaaacattctgtaagtca 180  
 |||||  
 QY 181 atcaataaaagaaaatctctgagaaaagaaaagaaatcaacgtataaataaaa 240  
 |||||  
 Db 181 atcaataaaagaaaatctctgagaaaagaaaagaaatcaacgtataaataaaa 240  
 |||||  
 QY 241 ggaattcaaaataaagatatcttgaagagaaaataagaatacatttacaagaacagcagag 300  
 |||||  
 Db 241 ggaattcaaaataaagatatcttgaagagaaaataagaatacatttacaagaacagcagag 300  
 |||||  
 QY 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaaggttctccagcag 360  
 |||||  
 Db 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaaggttctccagcag 360  
 |||||  
 QY 361 tctaaaagccaaacaaacaaagaaagcctaccttcttctgcaaaagcagctgacatg 420  
 |||||  
 Db 361 tctaaaagccaaacaaacaaagaaagcctaccttcttctgcaaaagcagctgacatg 420  
 |||||  
 QY 421 ggaaacttgaaagctatggagaaaatggctgacgctttgctatttggaaatttggcgctg 480  
 |||||  
 Db 421 ggaaacttgaaagctatggagaaaatggctgacgctttgctatttggaaatttggcgctg 480  
 |||||  
 QY 481 caaaataaaacagcagctatccaaattatagctccttggcttgaagaagatcatgtaa 540  
 |||||  
 Db 481 caaaataaaacagcagctatccaaattatagctccttggcttgaagaagatcatgtaa 540  
 |||||  
 QY 541 gcccaaaacgcattaggattttgtctcttgaataggaatggaatgatcatcaagct 600  
 |||||  
 Db 541 gcccaaaacgcattaggattttgtctcttgaataggaatggaatgatcatcaagct 600  
 |||||  
 QY 601 aaggcactgatattacacctttggaagctgctggagaaacatgatgtcccagattgatt 660  
 |||||  
 Db 601 aaggcactgatattacacctttggaagctgctggagaaacatgatgtcccagattgatt 660  
 |||||  
 QY 661 ttggggtacagatatgttcgggaatcaatgtctacagaattgtgaagtgtccctaaat 720  
 |||||  
 Db 661 ttggggtacagatatgttcgggaatcaatgtctacagaattgtgaagtgtccctaaat 720  
 |||||  
 QY 721 tattacaagaagtgccagattattgtctgacacatttgaaaaaagtgaaagtggttcca 780  
 |||||  
 Db 721 tattacaagaagtgccagattattgtctgacacatttgaaaaaagtgaaagtggttcca 780  
 |||||  
 QY 781 gtggaagaagtgaagactaacggaagacctgaaaaatctgagttctaacagtgagatttg 840



```
Db 781 gtggaaaagtggagtaacggaagacctgaaatctgagttctaaacagtgagatttg 840
Qy 841 gattggacataccaatactataaatttttggcagaaaggagagatttcagatacaa 900
Db 841 gattggacataccaatactataaatttttggcagaaaggagagatttcagatacaa 900
Qy 901 gtctctcttggaatactataatgtgagcagaaaggctctagatcaggattactacaaa 960
Db 901 gtctctcttggaatactataatgtgagcagaaaggctctagatcaggattactacaaa 960
Qy 961 gattacactacttcttaaaagcagaaaggccggggagtgcaaatgcaatgctcttcaag 1020
Db 961 gattacactacttcttaaaagcagaaaggccggggagtgcaaatgcaatgctcttcaag 1020
Qy 1021 gaaagatgtatttagagggaatgctgcgtgccgcaaaaataacgctactgctctcaag 1080
Db 1021 gaaagatgtatttagagggggaatgctgcgtgccgcaaaaataacgctactgctctcaag 1080
Qy 1081 tactttccatggcagccagtaaggcgaatgcaatgcgcttcaatggcttggtcttctt 1140
Db 1081 tactttccatggcagccagtaaggcgaatgcaatgcgcttcaatggcttggtcttctt 1140
Qy 1141 tactttcatggaaaaggagttcccttggaattatgcraagcaacttaatactttcagaaa 1200
Db 1141 tactttcatggaaaaggagttcccttggaattatgcraagcaacttaatactttcagaaa 1200
Qy 1201 gctgcgaaaaaagggtgcccgcagcacagttccagttcaggttagcttcatgtaactactctggc 1260
Db 1201 gctgcgaaaaaagggtgcccgcagcacagttccagttcaggttagcttcatgtaactactctggc 1260
Qy 1261 tctggaatatggaagattataaacttgcttcaaatatttttaactggcatctcagagt 1320
Db 1261 tctggaatatggaagattataaacttgcttcaaatatttttaactggcatctcagagt 1320
Qy 1321 gggcagccctgcgcaattattatctgccaagatgtatgcaacaggaacagagtagta 1380
Db 1321 gggcagccctgcgcaattattatctgccaagatgtatgcaacaggaacagagtagta 1380
Qy 1381 agatcatgcagaactgctgtggag 1404
Db 1381 agatcatgcagaactgctgtggag 1404

RESULT 5
AAD06380
ID AAD06380 standard; cDNA; 1728 BP.
XX
AC AAD06380;
XX
DT 10-AUG-2001 (first entry)
DE
DE Novel human protein (NHP) cDNA #7, sharing similarity with Notch ligands.
XX
KW Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
KW cholesterol metabolism; coronary artery disease; gene therapy;
KW cerebroprotective; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1728
FT /*tag= a
FT /product= "Novel human protein #7 (NHP)"
FT misc_feature 1177
FT /*tag= b
FT /note= "This ambiguity represents polymorphic site"
FT unsure 1177..1179
FT /*tag= c
```

```
FT note= "Encodes Glu"
PN WO200136636-A2.
XX
PD 25-MAY-2001.
XX
PF 16-NOV-2000; 2000WO-US31373.
XX
PR 17-NOV-1999; 99US-0165959.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
PI WPI; 2001-355635/37.
XX
DR P-PSDB; AAE02436.
XX
XX Novel isolated human polynucleotides encoding polypeptides sharing
PT sequence similarity with mammalian SEL-1 proteins which are negative
PT regulators of Notch family receptors, useful for treating diabetes,
PT cancer
XX
PS Disclosure; Page 37; 39pp; English.
XX
XX The present sequence is a cDNA encoding novel human protein (NHP) which
XX share structural similarity with animal Notch ligands, particularly
XX CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.
XX CC Notch receptors and their associated signalling pathways have been
XX CC associated with development, apoptosis, neuron growth and maintenance.
XX CC Labeled NHP probes can be used to screen a human genomic library which
XX CC is helpful for identifying polymorphisms, determining the genomic
XX CC structure of a given locus/allele and designing diagnostic tests. The
XX CC NHP is also useful in screening techniques for drugs which treat
XX CC symptomatic or phenotypic manifestations of perturbing the normal
XX CC function of NHP in the body. Nucleotide constructs encoding functional
XX CC NHPs, antisense, antisense molecules can be used in gene therapy or
XX CC approaches for modulating gene expression such as for preventing or
XX CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,
XX CC insulinomas), blood pressure abnormalities, neurodegenerative diseases
XX CC such as Parkinson's disease, stroke, vascular dementia and conditions
XX CC requiring modulation of fat and cholesterol metabolism such as coronary
XX CC artery disease.
XX
SQ Sequence 1728 BP; 592 A; 309 C; 380 G; 446 T; 1 other;

Query Match 67.8%; Score 1403.6; DB 22; Length 1728;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaagcccttgctctgttaatagagattgataattcttgggtgcacaattaaaact 60
Db 1 atgaagcccttgctctgttaatagagattgataattcttgggtgcacaattaaaact 60

Qy 61 atcaagcagaggaacataataaaagacaaaaggaagaaatgtcaccacacaggtatca 120
Db 61 atcaagcagaggaacataataaaagacaaaaggaagaaatgtcaccacacaggtatca 120

Qy 121 gtgaagaaatacaacaattatttaccacatatgttgaaacaaagaacatctagtatga 180
Db 121 gtgaagaaatacaacaattatttaccacatatgttgaaacaaagaacatctagtatga 180

Qy 181 atcaataaaagagaaaatctcctggagaaaagaagaatcaacgtaaaaataagataaaa 240
Db 181 atcaataaaagagaaaatctcctggagaaaagaagaatcaacgtaaaaataagataaaa 240

Qy 241 ggaattcaaaataaagatatcttgaagagaaataagaatacatttacaacaaagcagag 300
Db 241 ggaattcaaaataaagatatcttgaagagaaataagaatacatttacaacaaagcagag 300

Qy 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaagggttctccagcag 360
Db 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaagggttctccagcag 360
```

```
QY 361 tctaaaagccaaaacaaaagaaagcctacacctactcttttgcacaaagcagctgacatg 420
D 361 tctaaaagccaaaacaaaagaaagcctacacctactcttttgcacaaagcagctgacatg 420
QY 421 ggaacttgaaagctatgagagaaatggtgcagctgttgcctatttggaaatttggcgtg 480
D 421 ggaacttgaaagctatgagagaaatggtgcagctgttgcctatttggaaatttggcgtg 480
QY 481 caaataataacagcagctatccaattatatagctcttgccttaaaagaaggatcatgtaaa 540
D 481 caaataataacagcagctatccaattatatagctcttgccttaaaagaaggatcatgtaaa 540
QY 541 gccaaaacgcattaggatttgccttcttataagaaatggaatggaatataatatacaagt 600
D 541 gccaaaacgcattaggatttgccttcttataagaaatggaatggaatataatatacaagt 600
QY 601 aaggcactgatattacacaccttggaaagctgctgagaaacatgatgccagatgatt 660
D 601 aaggcactgatattacacaccttggaaagctgctgagaaacatgatgccagatgatt 660
QY 661 ttgggttacagatttgcgggaatacaattgttctacagaattgtgaagttgccttaagt 720
D 661 ttgggttacagatttgcgggaatacaattgttctacagaattgtgaagttgccttaagt 720
QY 721 tattacaagaaagtgcagatttatattgctgacacatttgaaaaagtgaaggtgttcca 780
D 721 tattacaagaaagtgcagatttatattgctgacacatttgaaaaagtgaaggtgttcca 780
QY 781 gtgaaaagtgcagactaacgaaagacactgaaatctgagttctaaacagtgagatttg 840
D 781 gtgaaaagtgcagactaacgaaagacactgaaatctgagttctaaacagtgagatttg 840
QY 841 gattggacataccaatactataaatttttggcagaagaggagattgttcagatacaa 900
D 841 gattggacataccaatactataaatttttggcagaagaggagattgttcagatacaa 900
QY 901 gtctcttggacaattacataatgtgaggaagaggtctagatcaggattactacaaa 960
D 901 gtctcttggacaattacataatgtgaggaagaggtctagatcaggattactacaaa 960
QY 961 gaattacacttcttaaagcagcaagcgcggagtgcaaatgccatggcatttata 1020
D 961 gaattacacttcttaaagcagcaagcgcggagtgcaaatgccatggcatttata 1020
QY 1021 ggaagatgtattagagggaatgctccgtgcccaaaataacgctactgcctcgaag 1080
D 1021 ggaagatgtattagagggaatgctccgtgcccaaaataacgctactgcctcgaag 1080
QY 1081 tactttccatggcagccagtaagggcaatgcaatcgcccttcattggccttctt 1140
D 1081 tactttccatggcagccagtaagggcaatgcaatcgcccttcattggccttctt 1140
QY 1141 tactttcattgaaaaggagttccctcgaattatgcraagcacttaatacttcagaaa 1200
D 1141 tactttcattgaaaaggagttccctcgaattatgcraagcacttaatacttcagaaa 1200
QY 1201 gctcgaaaaaggtgcccagcacagttccagttcagcttagccttcattactactcgc 1260
D 1201 gctcgaaaaaggtgcccagcacagttccagttcagcttagccttcattactactcgc 1260
QY 1261 tctggaatatgaaagattataaaacttgccttcaaatatttttacctggcattcagagt 1320
D 1261 tctggaatatgaaagattataaaacttgccttcaaatatttttacctggcattcagagt 1320
QY 1321 gggagccctgcatttattatgtgccaaagtgtatgcaacagggaacaggagtagta 1380
D 1321 gggagccctgcatttattatgtgccaaagtgtatgcaacagggaacaggagtagta 1380
QY 1381 agatcatgcagactgctgtgag 1404
D 1381 agatcatgcagactgctgtgag 1404
```

```
RESULT 6
AAD06379
ID AAD06379 standard; cDNA; 1731 BP.
XX
AC AAD06379;
XX
DT 10-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) cDNA #6, sharing similarity with Notch ligands.
XX
KW Human; nootropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
KW cholesterol metabolism; coronary artery disease; gene therapy;
KW cerebroprotective; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1731
FT /tag= a
FT /product= "Novel human protein #6 (NHP)"
FT misc_feature 1177
FT /tag= b
FT /note= "This ambiguity represents polymorphic site"
FT unsure 1177..1179
FT /tag= c
FT note= "Encodes Glu"
XX
PN WO200136636-A2.
XX
PD 25-MAY-2001.
XX
PF 16-NOV-2000; 2000WO-US31373.
XX
PR 17-NOV-1999; 99US-0165959.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
DR WPI; 2001-355635/37.
DR P-PSDB; AAE02435.
XX
PT Novel isolated human polynucleotides encoding polypeptides sharing
PT sequence similarity with mammalian SEL-1 proteins which are negative
PT regulators of Notch family receptors, useful for treating diabetes,
PT cancer
XX
PS Disclosure; Page 35-36; 39pp; English.
XX
CC The present sequence is a cDNA encoding novel human protein (NHP) which
CC share structural similarity with animal Notch ligands, particularly
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.
CC Notch receptors and their associated signalling pathways have been
CC associated with development, apoptosis, neuron growth and maintenance.
CC Labeled NHP probes can be used to screen a human genomic library which
CC is helpful for identifying polymorphisms, determining the genomic
CC structure of a given locus/allele and designing diagnostic tests. The
CC NHP is also useful in screening techniques for drugs which treats
CC symptomatic or phenotypic manifestations of perturbing the normal
CC function of NHP in the body. Nucleotide constructs encoding functional
CC NHPs, antisense, antisense molecules can be used in gene therapy
CC approaches for modulating gene expression such as for preventing or
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases
CC such as Parkinson's disease, stroke, vascular dementia and conditions
CC requiring modulation of fat and cholesterol metabolism such as coronary
CC artery disease.
XX
```



XX WPI; 2001-355635/37.  
DR P-PSDB; AAE02433.  
XX  
PT Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX  
PS Claim 5; Page 32; 39pp; English.  
XX  
CC The present sequence is a cDNA encoding novel human protein (NHP) which  
CC share structural similarity with animal Notch ligands, particularly  
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 1257 BP; 464 A; 206 C; 270 G; 316 T; 1 other;

Query Match 60.6%; Score 1254.6; DB 22; Length 1257;  
Best Local Similarity 100.08; Pred. No. 9.9e-292;  
Matches 1257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagccctgtctctgttaatagagattgataattcttgggtgcacaaattaaaact 60  
DB 1 atgaagccctgtctctgttaatagagattgataattcttgggtgcacaaattaaaact 60  
QY 61 atcaagcagaggaacataataaaagacaaagaaagaaatgtccacacacagtatca 120  
DB 61 atcaagcagaggaacataataaaagacaaagaaagaaatgtccacacacagtatca 120  
QY 121 gtgaacgaatacaacaattatttaccacatatgtggaacaaagaaacatctagtaatga 180  
DB 121 gtgaacgaatacaacaattatttaccacatatgtggaacaaagaaacatctagtaatga 180  
QY 181 atcaataaaagaaatactctctgtgagaaaaagaaagaaatcaacgtataaataaagaaaa 240  
DB 181 atcaataaaagaaatactctctgtgagaaaaagaaagaaatcaacgtataaataaagaaaa 240  
QY 241 ggaattcaaaataaagatatctgtgaagagaaataagaatcatttcaaaaagcaagcagag 300  
DB 241 ggaattcaaaataaagatatctgtgaagagaaataagaatcatttcaaaaagcaagcagag 300  
QY 301 aaaaattttacagatgaagagacagcattttaaagatgggcatcaaggttctccagcag 360  
DB 301 aaaaattttacagatgaagagacagcattttaaagatgggcatcaaggttctccagcag 360  
QY 361 tctaaaagccaaacaaacaaagaaagcctacctacttttggcacaagcagtcacatg 420  
DB 361 tctaaaagccaaacaaacaaagaaagcctacctacttttggcacaagcagtcacatg 420  
QY 421 ggaacttgaaagctatggagaaatgctgcagcgtttgtctatttggaaaattttggcgtg 480  
DB 421 ggaacttgaaagctatggagaaatgctgcagcgtttgtctatttggaaaattttggcgtg 480  
QY 481 caaataataacagcagctatcaattatatagctgcttggctaaagaggaatcatgtaaa 540  
DB 481 caaataataacagcagctatcaattatatagctgcttggctaaagaggaatcatgtaaa 540

QY 541 gcccaaacgcattaggtattttgtcttcttatggaatagggaatggaatgatcaagct 600  
DB 541 gcccaaacgcattaggtattttgtcttcttatggaatagggaatggaatgatcaagct 600  
QY 601 aaggcactgatatattacaccttttgaagtgctgagagaaacatgatgtcccagatgatt 660  
DB 601 aaggcactgatatattacaccttttgaagtgctgagagaaacatgatgtcccagatgatt 660  
QY 661 ttgggtgtacagattttgtcgggaataatgttctacagaattgtgaagttgcccctaagt 720  
DB 661 ttgggtgtacagattttgtcgggaataatgttctacagaattgtgaagttgcccctaagt 720  
QY 721 tattacaagaagtgagcagatttatattgtcacacatttgaaaaagtgaagtggttcca 780  
DB 721 tattacaagaagtgagcagatttatattgtcacacatttgaaaaagtgaagtggttcca 780  
QY 781 gtggaaaaagtgagactaacgaaagacctgaaatctgagttctaaacagtgagatttg 840  
DB 781 gtggaaaaagtgagactaacgaaagacctgaaatctgagttctaaacagtgagatttg 840  
QY 841 gattggacatatataactataataatttttgcagaaagagagagattgttcagatacaa 900  
DB 841 gattggacatatataactataataatttttgcagaaagagagagattgttcagatacaa 900  
QY 901 gtctctcttgacaaattacatttaatttggcagaaaggtcttagatcaggattactacaaa 960  
DB 901 gtctctcttgacaaattacatttaatttggcagaaaggtcttagatcaggattactacaaa 960  
QY 961 gcattacactacttcttaaaagcagaaagccggggagtgcataatgccatggcatttata 1020  
DB 961 gcattacactacttcttaaaagcagaaagccggggagtgcataatgccatggcatttata 1020  
QY 1021 ggaagagattatttagaggggaatgctccgtgcccgaataaaccgtactactcctcaag 1080  
DB 1021 ggaagagattatttagaggggaatgctccgtgcccgaataaaccgtactactcctcaag 1080  
QY 1081 tacttttccatggcagccagtaagggcaatgcaatcgcccttcattgggtgtcttctt 1140  
DB 1081 tacttttccatggcagccagtaagggcaatgcaatcgcccttcattgggtgtcttctt 1140  
QY 1141 tacttttccatggcagccagtaagggcaatgcaatcgcccttcattgggtgtcttctt 1200  
DB 1141 tacttttccatggcagccagtaagggcaatgcaatcgcccttcattgggtgtcttctt 1200  
QY 1201 gctgcgaaaaaggtggtccgcagcagcttccagttcaggttcttacttact 1255  
DB 1201 gctgcgaaaaaggtggtccgcagcagcttccagttcaggttcttacttact 1255

RESULT 8  
AA76578  
ID AA76578 standard; cDNA; 7885 BP.  
XX AA76578;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Human pancreas-specific tumour suppressor TSA305 gene SEQ ID NO:3.  
XX  
KW Human; pancreas; TSA305; pancreatic cancer; carcinoma; diagnosis;  
KW therapy; tumour suppressor; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 46..2430  
FT /\*tag= a  
XX  
PN WO9928457-A1.  
XX  
PD 10-JUN-1999.  
XX  
PF 25-NOV-1998; 98WO-JP05306.



[illegible]





Db 1254 gaacagtgaatgctagagaagatttgatttcaattattaccagtttcttagctgaaaaagg 1313  
QY 885 agatgttcagatacaagctctctcttggaacaattacatctaaattggcaggagaaggtctaga 944  
Db 1314 tgatgtacaagcacaggttggtcttggaacaactgcactgcagggcgtggagtaga 1373  
QY 945 tcaggattactacaagcattacactactcttctaaaggcagcaaaagccgggagtgcaaa 1004  
Db 1374 acagaatcatcagagagcatttgactacttcaatttagcagcaaaatgctggcaattcaca 1433  
QY 1005 tcccatgcatattataggaagatgtatttagagggaatgctgcogtccgcacaaataa 1064  
Db 1434 tgcattggccttttgggaagatttattcgaagaagtgacattgtacctcagagtaa 1493  
QY 1065 cgctactgccttcaagtacttttccatggcagccagtaaggaatgcaatcgcccttca 1124  
Db 1494 tgagacagctctccactactttaagaagctgctgacatgggcaaccacagttggacagag 1553  
QY 1125 tggccttggtctcttacttctaatggaagagtgctccctgaattatgccraagcact 1184  
Db 1554 tggccttggaatggcctactctctatggagaggagtgctcaagttaattatgatctagccct 1613  
QY 1185 taaactcttcgaagctgcgaaaaagggtggccgcagcacagttccagttaggctt 1244  
Db 1614 taagtattccggaagagctgcgaacaaggctgggtggatggcgagctacagcttggctc 1673  
QY 1245 catgtactacttggctctggaatggaaggattataaacttgccttcaaatatttta 1304  
Db 1674 catgtactaatggcattggagtcgaagagagattataaacaggcccttgaagtattttaa 1733  
QY 1305 cctggcatctcagatgggcagccctcctgccatttattatctggccaaagatgatgcaac 1364  
Db 1734 tttagctctcagggaggccatctctgttcttataacctagctcagatgcatgccag 1793  
QY 1365 aggaacagagtagtaagaatcagaaactcgtctggagctttataaaagggtctctgga 1424  
Db 1794 tggcacccgctgatgcgcatctgacacactcagtcggaggtgtgttaagaatgtatgta 1853  
QY 1425 actaggccactggctgagaatactctgcagcttacttgcctataagagtggtgatat 1484  
Db 1854 acgagccgttgcttgaaagcttatgactgcctataaacagctataaagatggcgatta 1913  
QY 1485 agattctctctgttcagatgcagctcttcagaaaaatggggtatgaagtgcctcaag 1544  
Db 1914 caatgctgcagtgatccagctactctctcgtgctgaacagggcgtatgaagtggcacaag 1973  
QY 1545 caattcagcatcatttggaaatcaaaaaggctaacattcttgaaaaagagaagatgta 1604  
Db 1974 caatgcagcttattctctgacagagaagcaagcatctgagtgagaatgaaactta 2033  
QY 1605 tccaatggcgtctctactatggaatcgagctgccattccaaggaatgcatttgcctagagt 1664  
Db 2034 tccagagcttctgctacattgaaacagggccgctcctaagcgtatactgtgctagaat 2093  
QY 1665 aaaaattgagattaccattactatgctatggctatgggactaagaagaactatacaaacgagc 1724  
Db 2094 taagctcgagactaccattctatgggttggcaccgagtgatagatttgaaactgcatt 2153  
QY 1725 cacacactacagcattgcagccaacaaataccacaacgcgaagcattgttcaatctggc 1784  
Db 2154 tattcaattaccgtctggctcttgagcagcaacacagtgcaagatgtgttcaatctggg 2213  
QY 1785 ttatatgtatgaacagcgtttaggcatcacaaaggacattcacttggccagagaattgta 1844  
Db 2214 atatatgcatgaaaggactgggcatataaacaggatattacccttgcgaacgcttttta 2273  
QY 1845 cgacatggtcgtcctaaacagagtcagatgcccacatacctgtgctcttcttgcgcgtcatga 1904  
Db 2274 tgacatggcagctgaagccagcccagatgcacaagttccagttcttccctagcctctgcaa 2333  
QY 1905 actggaatacagcatttgcgtccggatattcctgttttttaattcagttcacaacagatg 1964  
Db 2334 attggcgctcgtctatttcttgcagtcatacgg---gaaacaaacattcagatgtgtt 2390

QY 1965 gaactggctgaaactgggaacaacacacattggaccacactgggacttatttgtgattggcct 2024  
Db 2391 caccacacttgataggaccagcttttgggaacctgagtgaggaccttacctcatgacct 2450  
QY 2025 cattg 2029  
Db 2451 cattg 2455

RESULT 11  
AAAX77013  
ID AAX77013 standard; DNA; 3666 BP.  
XX  
AC AAX77013;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Mouse Sel-1L splice variant coding sequence.  
XX  
KW Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;  
KW insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;  
KW vascular dementia; Parkinson's disease; coronary heart disease;  
KW fat metabolism; cholesterol metabolism; ss.  
OS Mus sp.  
PN WO9927088-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 19-NOV-1998; 98WO-CA01058.  
XX  
PR 28-JUL-1998; 98US-0123549.  
PR 19-NOV-1997; 97US-0066140.  
XX  
PA (MOU ) MOUNT SINAI HOSPITAL.  
XX  
PI Bernstein A, Donoviel D;  
XX  
DR WPI; 1999-357833/30.  
DR P-PSDB; AAY18096.  
XX  
PT New Sel-1L nucleic acid molecule useful in the treatment of  
PT Alzheimer's disease, diabetes and cancer  
XX  
PS Claim 2; Page 71-73; 77pp; English.  
XX  
CC This sequence encodes a Sel-1L (Sel-1 like) protein (also previously  
CC known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and  
CC the host cell containing it can be used to prepare a Sel-1L protein.  
CC Compositions containing the Sel-1L proteins, or peptides that interfere  
CC with their binding can be used in a method for treating or preventing  
CC Alzheimer's disease, diabetes (especially insulin dependent diabetes  
CC mellitus), cancer (especially pancreatic cancer), stroke, vascular  
CC dementia, Parkinson's disease, or coronary heart disease. The  
CC compositions can also be used to treat conditions requiring modulation of  
CC fat or cholesterol metabolism.  
XX  
SQ Sequence 3666 BP; 956 A; 881 C; 953 G; 874 T; 2 other;

Query Match 28.8%; Score 596.4; DB 20; Length 3666;  
Best Local Similarity 60.0%; Pred. No. 1.6e-133;  
Matches 1031; Conservative 0; Mismatches 682; Indels 6; Gaps 2;  
QY 314 atgaaggagaccagctatttaagatgggcatacaagggttctccagcagctctaaagccaaa 373  
Db 455 aggaagcagagatgatctatcatcaggccgggataagatactgaatgaagcaataggaga 514  
QY 374 aacaaaaagaagaacccctacactcttcttgccaaagcagctgacatgggaactgaaag 433  
Db 515 gccaaaaagagagcatatcggttaccttcgaaggcagcgagcatgaatcacaccaag 574





PS Claim 2; Page 66-69; 77pp; English.

XX This sequence encodes a Sel-1L (Sel-1 like) protein (also previously known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and the host cell containing it can be used to prepare a Sel-1L protein.  
CC Compositions containing the Sel-1L proteins, or peptides that interfere with their binding can be used in a method for treating or preventing Alzheimer's disease, diabetes (especially insulin dependent diabetes mellitus), cancer (especially pancreatic cancer), stroke, vascular dementia, Parkinson's disease, or coronary heart disease. The compositions can also be used to treat conditions requiring modulation of fat or cholesterol metabolism.

XX Sequence 3667 BP; 956 A; 881 C; 954 G; 874 T; 2 other;

Query Match 28.7%; Score 594.8; DB 20; Length 3667;

Best Local Similarity 59.9%; Pred. No. 3.8e-133;

Matches 1030; Conservative 0; Mismatches 683; Indels 6; Gaps 2;

QY 314 atgaaggagaccagctatttaagatggccatcaaggcttcagcagctcaaaagccaaa 373

DB 455 aggaagcagagatgactatcagccggatgaagatactgaatggaagcaatagaaga 514

QY 374 acaaaaagaagaagcctacctactttttgccaagcagctgacatgggaacttgaag 433

DB 515 gccaaaagagagacatcggtaccttcagaagcagcagcagcgaatcacaccaaag 574

QY 434 ctatggagaaaatggctgacgtttgtctatttggaaaattttggcgtgcaaaataaagc 493

DB 575 cctggagagagtgctctatgctctcttctgttgggtattacctcacacagaataccagg 634

QY 494 cagctatccaatttatagctcttgctgaagaagcagctatgtaaaagcccaaaagcgc 553

DB 635 cagcacaagagatgtttgagaactgactgaggaagggctcccaaaagcagactggct 594

QY 554 taggattttgtctcttattgaataggaatggaatatgatcaagctaaagcagctgat 613

DB 695 ttggtttctacgctcttggtggcttgggttgaattcaagtcagcgaaggtcttgat 754

QY 614 attacaccttgaagtgctgagaaacatgatgtccagatgatttgggtacagat 673

DB 755 attactcttgagctcttgaggccaacctgatgacccatgatgtttgggtaccgct 814

QY 674 atttgcggaaatcaatgttctacagaattgtgaagtgcctcaagtattatcaagaag 733

DB 815 actgggtggcagtgagctccctcagagttgtgagtcggcactgaccatattcgtctg 874

QY 734 tggcagattatattgctgacacatttgaaaaaagtgaaagtggttccagtggaagagtg 793

DB 875 ttgccaatcatgttgctagtgatctctcctcaactggaggtctgtagtcagagaatc 934

QY 794 gactaacggaagacctgaaatctgagttctaacagtgagatttggattgggacat 853

DB 935 ggtcgcctgatgaagtggaaaaccggggtgaacagtggggtctgggaagacctga 994

QY 854 accaactactataaattttggcagaaaagagagatgttcagatacaagtctctctggac 913

DB 995 ttacagtattacagttctcagtgagaagggtagctccaaacacaggttggcctgggac 1054

QY 914 aattacatctaattggcaggaaggtcctagatcaggattactacaagaacattacact 973

DB 1055 agctgcatctgcatggagcgtggagtgaagaagaataccacagagagcgtttgactact 1114

QY 974 tcttaaaagcagcaagccggagtgcaaatgccatggcatttatagaaagatgatt 1033

DB 1115 tcaacttagcagcaaatgctggaatctcaatgctatggcctctcctgggaagatgatt 1174

QY 1034 tagagggaaatgctgctgcgcgcaaaaataacgctactgcttccactactttccatgg 1093

DB 1175 ctgaggaagatgacatcgtaacctgacagagtaatgagacgcacttcaactacttaagaa 1234

QY 1094 cagccagtaagggaatgcaatcgcccttcattggccttggtgtcttcttacttcaatg 1153

DB 1235 ctgctgacatggcgcaaccocgtgggacagagcggttggaatggcctacctctacggaa 1294

QY 1154 aaggagttcccttgatattatgaccraagcaacttaatactttcagaaaaagctgggaaaaag 1213

DB 1295 gaggcgttcaagttaattatgacctggcctcaagtattccagaaagctgctgagcaag 1354

QY 1214 ggtggccgcagcacagctccagtttaggttcatgtactactctggtctggaatatgga 1273

DB 1355 gctgggtggacgggcagctgcagctcgctctatgtactacaatggcattggagtgcaaga 1414

QY 1274 aggtataaaacttgccttcaaatatttttacctggcactctcagagtgggcagccctgg 1333

DB 1415 gagattaaagcagccttggaagtattttaactgtggtcttcaaggagccatattctgg 1474

QY 1334 ccattttattatctggc---caagatgtatgcaacaggaacagagtagtaagaatcagca 1390

DB 1475 cttttataacctgcacnnaagatgcagcgcagcgccaggggtgtagcgctctgctgc 1534

QY 1391 gaactgctgtggagctttataaaggtgctgtgaactaggccactgggctggagaaattcc 1450

DB 1535 acactgcagtgaggtgtttaagaatgtgtgagcgaggtcgctggctcagagagactga 1594

QY 1451 tgacagcttactttgcttataaggatggtgatatagattcttctcttctgttcagtgatgc 1510

DB 1595 tgactgctcacacagctataaggatgagactacaatgctgcagtggtccagtaacctcc 1654

QY 1511 tgccttcagaaaatgggttatgaatgactcaagcaaatccagcattcatttggaaatcta 1570

DB 1655 tgctgctgagcagcgtcacggtggcgagcaagcagccttcaacctcgaccaga 1714

QY 1571 aaaaagcttaacattctgaaaagagaagatgtatccaatggcgtcttctctatggaatc 1630

DB 1715 gagaagcaacctatgagtgagaatgaaacttaccacagactttactgcatggagaaca 1774

QY 1631 gagctgccattcaagcgaatgcatgttgtagagtgaaataattggagattaccattactatg 1690

DB 1775 gggcgcctcccaaggttacactgtggctagaattaaagcttgagactaccactctctatg 1834

QY 1691 gctatggagactaagaagactatcaaacagcagccacacacacacagcattgcagcaaca 1750

DB 1835 gcttggcactgatgtggattatgagccgcatattattaccgctggtctctgagc 1894

QY 1751 aataccacaacgcgaagcctatccaatctggtcttatatgtatgaacacgcttaggca 1810

DB 1895 agcagcacagcgcccaagctatgtttaacctgggtctacatgcacgagaagggctaggca 1954

QY 1811 tcacaaaaggacattcacttggccagaagattgtacgacatgctgctcaaaacagctccag 1870

DB 1955 ttaacagagacattcaccttgcaaacgcgttttatgacatggcagccggaagctagccag 2014

QY 1871 atgcccacatacctgtgctcttgcgttcacatgaaactggaaactacgcaatttgcctcggg 1930

DB 2015 atgcacaagtaacctgtgttccctgcactctgcaaatagggtgtgctctattcttacctg 2074

QY 1931 atactcgttttttaactagttccaacagagatggaactggcgtggaactggcaactggcaacacca 1990

DB 2075 acatacgggaagcaa---acattcgagatctattcacacaactgatatggaccagcttt 2131

QY 1991 ttggaccacactgggacttatttggattggcctcattg 2029

DB 2132 tgggaccggagtgggacctttaccctcatgaccatcattg 2170

RESULT 13

AAF93617

ID AAF93617 standard; cDNA; 557 BP.

XX AAF93617;

XX AAF93617;

DT 21-MAY-2001 (first entry)

XX cDNA encoding SRT protein isolated from testis tissue SEQ ID 438.





Query Match									
Best Local Similarity		21.1%		Score 437.6		DB 20		Length 2109	
Matches 690		Conservative		0		Mismatches 399		Indels 3	
Gaps									
Qy	938	gtctagatcaggattactacaaagcaattacactactcttcttaaaagcgacgaagccggga	997						
Db	67	gtcgacacagatcatcagagagcaatttgactactctcaatttagcagcaaatgctggca	126						
Qy	998	gtgcaaatgcctatggcattataggaagatgtatttagagggaatctgcctgcgcg	1057						
Db	127	attcaatgcctggccttttgggaaagatgtatctggagggaagtgacatttaacct	186						
Qy	1058	aaaataacgctactgccttcaagctcttccatgcgcagcagtaaggcgcaatgcaatg	1117						
Db	187	agagtaatgagacgctctccactactttaagaaagctgctgacatgggcaaccagttg	246						
Qy	1118	gccttcaatgggctgttcttcttacttctcattggaagaggaggttcccctgaattatg	1177						
Db	247	gacagatgggcttggaatggcctacctctatgggagaggagttcaagttaattatg	306						
Qy	1178	aagcacttaacttctgaagctgcgcgaaaaagggtgccgcagcgcacagttccagt	1237						
Db	307	tagcccttaagtattccagaaagctgcgcagacaggctgggtggatgggcagctacgc	366						
Qy	1238	tagcttcatgtactactctgctctggaatatgaaaggattataaacttgccttcaat	1297						
Db	367	ttggtccatgtactataatgcatggagttgagctcaagagagattataaacagccttga	426						
Qy	1298	attttaccctggcatctcagatggtggcagccctgccttattattctggtggcgaagt	1357						
Db	427	attttaatttagcttctcaggaggccatatcttgcttctctataacctagctcagatgc	486						
Qy	1358	atgcacaggaacagagtagtaagaatcatgcagaaacgtcgtggagctttataaaggtg	1417						
Db	487	atgcccagtggcacccggctgctgcctatgtctcacctgcagtggtgtttaagaagt	546						
Qy	1418	ctctgaactaggccactggcttgagaaattcctgacagcttaacttgcctataaggtg	1477						
Db	547	tatgtgaacgagccgttggtctgaaagcttatgactgcctataacagctataaagatg	606						
Qy	1478	gtgatatagattcttcttgttcagtatgctactgcttgcagaaatgggtgatgaatg	1537						
Db	607	gcgattacaatgctgcagtgatccagctacctctctgctggaacaggctatgaatgg	666						
Qy	1538	ctcaaaagcaattcagcatctatttggaaattcaaaaagcttaacaattcttgaaaaagag	1597						
Db	667	cacaaagcaatgcagccttattctgtatcagagagaagcagatgttagtgagaatg	726						
Qy	1598	agatgtaccaatggcgtctctctatgaaatcgagctgacctcaaggccaatgcattg	1657						
Db	727	aaacttatccagagcttgctacattggaaacagggcgccctctcaaggctatactgtg	786						
Qy	1658	ctagagtataaataatggagattaccattactatggcctatgggactaagaagactatcaa	1717						
Db	787	ctagaattaaagctcggagactaacattctatggtttggttggcacccgatgtagattgaa	846						
Qy	1718	cagcagccacactacagcatttgacgcaacaataccacaacggcgcaagcgaatgttca	1777						
Db	847	ctgcatttatccatccgtctggtctctgacgacacacagctgcaacgctatgttta	906						
Qy	1778	atctggcttatatgatcaacggcttaggcatacacaaaggacattcaacttggccagaa	1837						
Db	907	atctgggatataatgatgagaaggactgggcatttaacacaggattatccacttgcgaac	966						
Qy	1838	gatttgtacgacatggctgctcaaacaggtccagatgcccacatacctgtgctctttccg	1897						
Db	967	gtttttatgacatggcagctggaagccagcccgatgcaaatgtccagttctcttagccc	1026						
Qy	1898	tcattgaactggaaactagcatttgcctccgggatatctgttttttaactcagttcaaa	1957						
Db	1027	tcctgcaaatgggctgctctattcttcgtacgtacatacagg---gaaacaacacttcag	1083						

	Qy	1958	cgaatggaactggctgaaacttgacacacaccattgagaccacactggagacctattttgtga	2017
	Dd	1084	atatgtcaccccaactgataggacagccttcttgggacctgagtgggaccttatcctcca	1143
	Qy	2018	ttaggccttcattg	2029
	Dd	1144	tgaccatcatctg	1155

Search completed: April 8, 2002, 21:34:51  
Job time: 3162 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2002, 20:24:18 ; Search time 1193 Seconds  
(without alignments)  
18645.232 Million cell updates/sec

Title: US-09-714-882-1

Perfect score: 2070

Sequence: 1 atgaagccctgtctctgtt.....ttagaatcaccatgggtag 2070

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estro:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	608.8	29.4	3055	12 AK005023	AK005023 Mus muscu
2	304.8	14.7	421	11 BF153071	BF153071 uz99g07.y
3	252.8	12.2	630	11 BI067239	BI067239 pgfln.pk0
4	243.8	11.8	592	10 BF680423	BF680423 df79c09.y
5	220.2	10.6	343	11 BF509718	BF509718 UI-H-Bf4-
6	215.2	10.4	549	11 BG577404	BG577404 N115 SSH-
7	198.2	9.6	459	10 AA310028	AA310028 EST180874
8	196.6	9.5	694	10 AL596660	AL596660 DKZp451L
9	195	9.4	632	11 BE912256	BE912256 G01665066
10	170.8	8.3	932	10 BE570424	BE570424 G01333322
11	168.6	8.1	492	11 BE995163	BE995163 UI-M-CG0P
12	166.6	8.0	659	11 BF727590	BF727590 SMOV3MCAM

13	165.8	8.0	404	11 BF359310	BF359310 RCS-ET008
14	165.6	8.0	721	11 BG197665	BG197665 RST16894
15	165.6	8.0	763	11 BG184301	BG184301 RST3225 A
16	163	7.9	553	10 AL596788	AL596788 DKFZp451J
17	157.6	7.6	776	11 BG185764	BG185764 RST4716 A
18	154.6	7.5	366	11 BG188306	BG188306 RST7323 A
19	153.8	7.4	179	11 BG182541	BG182541 RST1417 A
20	151.6	7.3	716	13 AQ485268	AQ485268 RST11-2
21	150.6	7.3	494	10 AL589253	AL589253 DKFZp451C
22	150.2	7.3	428	10 AL751663	AL751663 cn11c01.x
23	147.8	7.1	424	11 BG191061	BG191061 RST10031
24	147.2	7.1	464	11 BG201173	BG201173 RST20501
25	143.8	6.9	542	10 AL589397	AL589397 DKFZp451N
26	141.8	6.9	178	11 BG216522	BG216522 RST36102
27	141.6	6.8	174	11 BG195560	BG195560 RST14753
28	141	6.8	564	13 AQ984054	AQ984054 RST23-3
29	140.8	6.8	353	11 BG184754	BG184754 RST3576 A
30	140.4	6.8	490	10 AI425637	AI425637 md58h04.y
31	140.2	6.8	177	11 BG204271	BG204271 RST23672
32	138.8	6.7	534	11 BG880030	BG880030 ib73h09.y
33	137	6.6	174	11 BG203754	BG203754 RST23129
34	137	6.6	344	11 W62650	W62650 md58h04.r1
35	135	6.5	622	11 BF501529	BF501529 AT16891.5
36	134.8	6.5	178	11 BG197561	BG197561 RST16806
37	134.4	6.5	595	10 AW281997	AW281997 fj60e07.x
38	133.8	6.5	362	11 BG218395	BG218395 RST38263
39	132.2	6.4	711	10 AV712962	AV712962 AV712962
40	131.8	6.4	573	10 AI552501	AI552501 m194h06.y
41	130.6	6.3	176	11 BG219992	BG219992 RST39784
42	130.6	6.3	591	10 AA288149	AA288149 vbl5e09.r
43	128	6.2	212	11 BG184106	BG184106 RST3024 A
44	124.4	6.0	943	11 BE912650	BE912650 G01663693
45	119.8	5.8	270	10 AV265793	AV265793 AV265793

## ALIGNMENTS

RESULT	1
AK005023	
LOCUS	AK005023 3055 bp mRNA HTC 05-JUL-2001
DEFINITION	Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300016D21, full insert sequence.
ACCESSION	AK005023
VERSION	AK005023.1 GI:12836652
KEYWORDS	CAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 3055)
TITLE	Carninci, P. and Hayashizaki, Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Methods in enzymology. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	2 (bases 1 to 3055)
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE	Genome research. 10 (10), 1617-1630 (2000)
PUBMED	20499374
AUTHORS	3 (bases 1 to 3055)
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome research. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861  
 4 (bases 1 to 3055)  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5 (bases 1 to 3055)  
 Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,  
 Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,  
 Hanagaka,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,  
 Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,  
 Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,  
 Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,  
 Okido,T., Owa,C., Saito,H., Saito,R., Sakai,K., Sano,H.,  
 Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,  
 Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
 Tanaka,T., Teijima,Y., Toya,T., Yamamura,T., Yasunishi,A.,  
 Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
 Direct Submission  
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp.  
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
 Fax:81-45-503-9216)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5', GAGAGAGAGCGCGCACTCGAGTGTGTTTTTTTATVN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. Second  
 strand cDNA was prepared with the primer adapter of sequence[5'  
 GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATTAACCGCCCCC 3']. cDNA was  
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
 XhoI. Host: SOLR.  
 Location/Qualifiers  
 1..3055  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /db\_xref="MGI:1329016"  
 /db\_xref="MGI:1907284"  
 /clone="J300016D21"  
 /sex="male"  
 /tissue\_type="liver"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 99..2471  
 /note="putative"  
 /codon\_start=1  
 /protein\_id="BAB23750.1"  
 /db\_xref="GI:12836653"  
 /translation="MOVRYRLSLLLLCVLLGSAATSDDKTNQDLSQSKSLPTDDE  
 SVKDHTTCKVVGQIFVDSERAEVSLQDEDSKTKQEEISFLESPNPSKTYEE  
 LKRVAPVLAETGHAEGPCHFPFLDKREKTSQDGRGLMCATTYDYKTDK  
 WGFETDEDAKRRQAEAMTYQAGMKLINGSNRKSEAREYRLQKAGSMHNTKL  
 ERYVALLCEDYLTQIAKAEKFLTEBEGSPKGTGLFGLVAGLVNSAKALV  
 YTFGALGNTLHMLTYRYKAWFAGIGVLCESALTHYRLVANHVASDILSTGGSVQ  
 RRLPDEVENPGMSGMLIEDLLQYQYFLAEKGDVQAVCLQLHLHGGRVQNHQR  
 ARDYFNLANAGNSHMAFLKGMVSEGSIVDPQSNETHYFRKADRVPGVQSGIG  
 MAYLYRGVQVNYDLALKYFQKAAEQGVQDQGLGSMYINGVGRKDYKQALKYFNL  
 ASQGVHLYFNLAHQAGSGVYMRSCHTDAVELFNKVCERGRWSRLMTAYNKDSD  
 YNQAQVYLLLAQQEYVAGSAAFTLDQREATIVGNETIPRALLHWNRAASQGVTY  
 YNQAQVYLLLAQQEYVAGSAAFTLDQREATIVGNETIPRALLHWNRAASQGVTY

BASE COUNT	788 a	735 c	830 g	702 t	Query Match	29.4%;	Score 608.8;	DB 12;	Length 3055;
ORIGIN					Best Local Similarity	60.1%;	Pred. No. 3.9e-102;		
					Matches 1031;	Conservative	0;	Mismatches 682;	Indels 3; Gaps 1;
QY	314	atgaagagaccagctatttaagatggcgatcaaggtctccagcagctctcaaaagccaaa	373						
Db	628	AGGAAGCAGAGATGATCTATCAGGCCGGGATGAAGATACTCAATATGGAAGCAATAGGAAGA	687						
QY	374	aacaaaaagagaagcctaccactctttttccaaagcagctgacatgggaaccttgaag	433						
Db	688	GCCAAAAGAGAGAGCATATCGGTACCTTCAGAAAGCCAGCAGGCATCAATACACACCAAG	747						
QY	434	ctatggagaaaaatggctgacgcttttgctatttggaaaattttggcgtgcaaaataataacag	493						
Db	748	CCCTGGAGAGAGTCTCTATGCTCTCTTGTGTGATTACCTCACACAGAATATCCAGG	807						
QY	494	cagctatcccaattatatgagcttccttggctaaagaagatcatgtaaagcccaaaacgcat	553						
Db	808	CAGCCAAAGAGATGTTTGAAGAACTGACTCAGGAAGGGTCTCCAAAGGACAGACTGGTC	867						
QY	554	taggatttttgccttcttattggaataggaaatggaaatatgataagctaaagcactgat	613						
Db	868	TTGGCTTTCTCTACGGCTTCTGGGCTTGGTGTTAATCAAGTCAGGCCAAAGGCTTGTGAT	927						
QY	614	attacaccttggaaagtgcggaggaaaacatgatgtccacagatgattttgggggtacagat	673						
Db	928	ATTATACTTTTCGAGCTCTTGGAGGCAACCTGATAGCCCATATGATTTTGGGTACCGCT	987						
QY	674	atttgtcgggaatcaaatgtttctcacaaattgtgaagtgccttaagtattacaagaag	733						
Db	988	ACTGGGCTGGCATCGGAGTCTCCAGAGTTGTGAGTCGGCACTCACCCATATATCGTCTTG	1047						
QY	734	tggcagattattgtcgcacacatttgaaaaaagtgaaagtgcttcagtggaagaaagtga	793						
Db	1048	TTGCCAATCATGTGTGCTAGTATCTCCCTAACTGGAGGCTCTGTAGTCCACAGAGATAC	1107						
QY	794	gactaacggaaaagcctgaaaaatctgagttcctaacagtgagatttggattgggacatat	853						
Db	1108	GGCTGCCGATGAAGTGGAAAAACCCGGGGATGACACAGTGGGATGCTGGAGAGAGACTGA	1167						
QY	854	accaatactataaatttttggcgaagaaggagatgttcagatatacagttctctcttggac	913						
Db	1168	TTCAGTATTACCAAGTTCTTACGTAGTGAGAAGGTCAGCTCCAAAGCACAGGTTGGCTCGGAC	1227						
QY	914	aattacatctaatttggcagaaaggcttagatcaggattactacaagcatatacactact	973						
Db	1228	AGCTGCATCTGCATGGAGGCGTGGAGTAGAAGACAGAATACACAGAGAGCGCTTTGACTACT	1287						
QY	974	tcttaaggcagcaaaagccggcggagtgcgaatgccatggcattataggaagaatgatt	1033						
Db	1288	TCAACTTAGCAGCAAAATGCTGGCAATTACATGCTATGGCCTTCCCTGGGAAAGATGTATT	1347						
QY	1034	tagaggggaatgctgcggtgccgcaaaataacgctactgccttcaagctactttccatgg	1093						
Db	1348	CTGAAGGAAGTGACATCGTACCTCAGAGTAAATGAGAGGGCACATTCTACTATTAAAGAAAG	1407						
QY	1094	cagccagtaagggcaatgcbaatcggccttcattggcgttggtcttcttactttcatggaa	1153						
Db	1408	CTGCTGACATATGGCAACCCCGCTGGGACAGACGGGGCTTGGAAATGGCCTACTCTACGGAA	1467						
QY	1154	aaggagttccctgaattatgtccraagcacttaaaacttccagaagctcgcgaaaaag	1213						
Db	1468	GAGCGTTCAAGTTAAATATGACCTGGCCCTCAAGTATTTCCAGAAAGCTGCTGAGCAAG	1527						
QY	1214	ggtggccgcgacgacagttccagttcaggttcattgactactctggtccttggaatatgga	1273						



Db	1528	CTGTGGTGGACGGGACGTGCAGCTCGGCTCTATCTACTACAATGGCATTTGGAGTCAAGA	158
QY	1274	aggattataaacttgccttccaaatatattttacctggcatctcagatgggcagccctcg	1333
Db	1588	GAGATTATTAAGCAGGCGCTTGAAGATATTTAAATCTGGCTTCTCAAGGAGGCCATATCTTGG	1647
QY	1334	ceatttattatctgcgcaagatgtatcaacaggaacaggagtagtaagatcatgcagaa	1393
Db	1648	CTTCTATTAACCTTCGCACAGATGCACGCCAGCGGCACAGGGGTGATCGGGTCTGTGCACA	1707
QY	1394	ctgctgtggagctttataaagtgctgtgtaactaggccactgggctggagaaattctctga	1453
Db	1708	CTGCAGTGGAGTGTTTAGATATGTCTGTGACGGAGGTCTGCTGCAGAGAGATGATGA	1767
QY	1454	cagcttacttgcctataaaggatggtgatatagattcttctctgttctcagtatgactgc	1513
Db	1768	CTGCCTFACACACAGCTATAAGGATGAGGACTACAAATGCTGCAGTGGTCCAGTACCTCTGC	1827
QY	1514	tgcagaataatggggtatgaagtgcctcaagcaatcagcattcatitttggaatctaaaa	1573
Db	1828	TGGCTCAGCAGGCGCTACGAGGTGGCGCAGAGCAACGACGCTTCATCTCCACCAGAGAG	1887
QY	1574	aggctaacattcttgaaaaagagaagtatacccaatggcgtctctctatggaaatcgag	1633
Db	1888	AAGCAACCAATTGTAGTGTAGATGAACCTTACCCACGAGCTTTACTGCATTTGGACAGGG	1947
QY	1634	ctgccattcaaggcaatgcatttgcctadagtaaaaaattggagattaccattactatggct	1693
Db	1948	CCGCCCTCCCAAGGTTACACTGTGGCTAGAAATTAAGCTTTGGAGACTACCACATCTTATGGCT	2007
QY	1694	atgggactaagaagaactatacaacagcagcgcacacactcagcattgcagcccacaat	1753
Db	2008	TTGGCACTGATGTGGATTATGACCGCATTTATTTCATTACCGCTTGCTTCTGAGCAGC	2067
QY	1754	accacaacgcgaagccatttcactctggcttatgtatgtacacagcgtctaggcatca	1813
Db	2068	AGCACAGCGCCCAAGCATGTTTAACTTGGGCTACATGCACGAGAAGGCGCTAGGCATTA	2127
QY	1814	caaagagacattcaacttggccagaagattgtacgacatggctgtcctcaaacagatccagatg	1873
Db	2128	AACAGAGACATTCACCTTGCAAAACGCTTTATAGATGGCAGCGCAAGCTAGCCCCAGATG	2187
QY	1874	ccacataactgtgctcttgcgcgcatgaaactgaaactacgatttgcctcgggata	1933
Db	2188	CACAAGTACCTGTGTCTCGCACTCTGCAAATTTAGGTGTCGTCTATTTTCTTTACAGTACA	2247
QY	1934	tcctgttttttaatcagttccacaacgagatggaactggctgaaactggcagaacaccattg	1993
Db	2248	TACGGGAAGCAA--ACATTTCGAGATCTATTTCACACAACCTGGATATGGACCGCTTTTGG	2304
QY	1994	gaccacactgggaactatttgtattggcctcattg	2029
Db	2305	GACCCGAGTGGGACCTTTTACCTCATGACCATCATTTG	2340

RESULT	2	
BF153071		
LOCUS	421 bp	mRNA
DEFINITION	BF153071.1 421 bp EST 26-OCT-2000	
	uz19907.y1 NCI-CGAP_Tel Mus musculus cDNA clone IMAGE:3809845 5'	
	similar to TC:Q9UGD3 Q9UGD3 DJ84266.2 ; mRNA sequence.	
ACCESSION	BF153071	
VERSION	BF153071.1	GI:11034466
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 421)	
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
	Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D.	

**ORGANISM**

Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 630)  
Cogburn, L.A., Morgan, R.W. and Burnside, J.  
Chicken ESTs from fat  
Unpublished (2001)  
**TITLE**  
Contact: Larry A. Cogburn  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1335  
Fax: 302-831-2822  
Email: cogburn@udel.edu, www.chickest.udel.edu.

**REFERENCE**

**AUTHORS**  
COGURN, L.A., MORGAN, R.W. AND BURNSIDE, J.  
CHICKEN ESTS FROM FAT  
UNPUBLISHED (2001)

**JOURNAL**

**COMMENT**

**FEATURES**

source  
1..630  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/clone="pgfin.pk011.e10"  
/clone\_lib="normalized chicken fat cDNA library"  
/sex="Male and Female"  
/tissue\_type="fat"  
/lab\_host="E.coli EMPH10B"  
/note="Vector: pSPORL1"

**BASE COUNT** 183 a 115 c 146 g 181 t 5 others

**ORIGIN**

Query Match 12.2%; Score 252.8; DB 11; Length 630;  
Best Local Similarity 66.5%; Pred. No. 8.8e-37;  
Matches 362; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 1393 actgctgtggagcctttatagaaggctgttgtaactagccactggctggcgagaaatttcctg 1452  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 13 ACTGCTGTTCAGTGGTTAAGATGTTTGTAACGTGCAGCGCTGCTCAAGAAGACTTATG 72  
QY 1453 acagcttaaccttgcctataagagtgtgatagattctcttgttcagtagtcactg 1512  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 73 ACTGCCATAAACAGCTATAAAGATGCGATTCAAATCTGCTGTGTTGTCAGTATCTTCTT 132  
QY 1513 ctgtcgaagaatgggtatgaactcaaacgaattcagcatctatttggaaatctaaa 1572  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 133 TTGGCAGACAAGGCTATGAAGTTGCACAAGCAATGCTGCTTTCTATCTACTTGTATCAAAAA 192  
QY 1573 aagcctaaccattcttgaagaagaagatgtatccaatggcgctctctctaatggaatcga 1632  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 193 GAAGCTAGCATAGTAGAGAAAATGAACATATCTCCGAGCATCTTCATCGGAATAGA 252  
QY 1633 gctgcattcaaggaatgcatttgcctagagtaaataatggagattacattactatggc 1692  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 253 GCAGCTTCTCAAGGATATACAGTCAAGAATAAACTGGAGATTACCATTTCTATGGA 312  
QY 1693 tatggactaagaagactatacaacagcagccacaactacagcattgcagcaacaaa 1752  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 313 TTTGGTACTGATGTTGATTATGAACACAGCATTTATTCATCTATCCAGCTGGCATCAGACAA 372  
QY 1753 taccacaacgcgaagccattgttcaatctggcttatgtatgaacacgcgcttaggcac 1812  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 373 CAGCACAGTGCATCAGCATGTTTAACTGGCGCTACATGCATGAGAGGGATGCGGCATC 432  
QY 1813 acaaaaggacattcaacttggccagaagattgttacgacatggctgctcaaacgagtcacag 1872  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 433 AAGCAGGATATTATCTTTGCAAAACGATTCATGATATGGCTGTGCTGAAGCAAGCCCAGAT 492  
QY 1873 gccacataacctgtctcttttgcgctcatgaactggaactacgcatttgcctcgggat 1932  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 493 GCCACAGGTTCCAAGTCTTCTAGCTCTTTTGAAGCTTTGGAGTGATATAGCTGTGCGATAT 552  
QY 1933 atcc 1936  
||| |  
Db 553 ATAC 556

---

**RESULT** 4

BE680423 592 bp mRNA EST 18-APR-2001  
LOCUS df79c09.y1 Xenopus laevis oocyte non normalized Xenopus laevis cdna  
DEFINITION clone IMAGE:3745552 5' similar to TR:Q92266 Q92266 SELLL. ; mRNA sequence.

ACCESION BE680423 GI:10064262  
VERSION BE680423  
KEYWORDS African clawed frog.  
SOURCE Xenopus laevis  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus.  
1 (bases 1 to 592)  
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,  
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,  
B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,  
Waterston,R. and Wilson,R.  
WashU Xenopus EST project, 1999  
Unpublished (1999)  
Other ESTs: df79c09.xl  
Contact: Sandy Clifton, Ph.D.  
WashU Xenopus EST project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library constructed by Bruce Blumberg  
Library normalized by Jihwan Song  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: Xenopus clone distribution information for  
this library can be found through Research Genetics, visit their  
web page at: http://www.resgen.com/ Please reference the id listed  
below when ordering this clone: Source lab clone id - xlnnoc002f17  
Trace assembled overall poor quality  
Seq primer: -40RP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1..592  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="IMAGE:3745552"  
/clone\_lib="Xenopus laevis oocyte non normalized"  
/tissue\_type="oocyte (stages 5 and 6)"  
/lab\_host="Top-10 F"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was prepared from 2ug of poly A+ RNA.  
EcoRI-XhoI cut cDNA was then ligated into UniZap-XR  
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'  
end. SS-library phagemids were prepared by mass excision  
from the original library and normalized by hybridization  
to biotinylated driver (prepared from the same library by  
PCR) to Cot-omega of 11. After removal of hybrids and  
excess driver by streptavidin sepharose chromatography,  
the ss-phagemids were made double stranded and  
electroporated into Top-10 F'. Original library  
construction by Bruce Blumberg (Blumberg et al., 1991  
Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,  
2923-2935). Note: This is a Xenopus Gene Collection (XGC)  
library."

**BASE COUNT** 161 a 143 c 142 g 146 t

**ORIGIN**

Query Match 11.8%; Score 243.8; DB 10; Length 592;  
Best Local Similarity 63.3%; Pred. No. 4e-35;  
Matches 374; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 1275 ggattataaacttgcttcaaataattttacctgcgcatctcagatggcgagccccctgcg 1334  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 GGAATTCCCAACAGCCCTGAATATTTTAACTGGCTGCATCACAAGGGGGGCACATTTTGGC 60









Db 358 CTTAGCATTAACAGGACATTCACCTTGCACAAACGCTTTTATGACATGCGCCGCAAGC 417  
 QY 1863 gagtcagatgccacacatctgtctcttctgctgcatgaaactggaactacacattt 1922  
 Db 418 TAGCCAGATGACAGACAGTACCTGTCTCCGACACTCTGCAAAATAGGTGTGCTATT 477  
 QY 1923 gctccgggatctctgttttttaacagttcacacagagatggaactggctgaaactgga 1982  
 Db 478 CTTACAGTACATACGGGAAGCAA---ACATTCGAGATCTATTACACAACTGGATATGGA 534  
 QY 1983 caacaccattggacacacactgggaacttattgtgattggcctcattgttctt 2034  
 Db 535 CCAGCTTTTGGGACCGAGTGGGACCTTTTACCTCATGACCATCATTTGACCGT 586

## RESULT 11

BE995163/c 492 bp mRNA EST 05-OCT-2000  
 LOCUS UI-M-CG0p-b11-h-01-0-UI.s1 NIH\_BMAP\_Ret4\_S2 Mus musculus cDNA clone  
 DEFINITION UI-M-CG0p-b11-h-01-0-UI 3', mRNA sequence.

ACCESSION BE995163  
 VERSION BE995163.1 GI:10679134  
 KEYWORDS EST.

SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 492)  
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704477  
 COMMENT Contact: Chin, H

National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mEST@mail.nih.gov

Oligo-dT track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA library Preparation: M.B.  
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements  
 Seq primer: M13 Forward  
 POLYA=No.

## FEATURES

source

Location/Qualifiers

1..492  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-CG0p-b11-h-01-0-UI"  
 /clone\_lib="NIH\_BMAP\_Ret4\_S2"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT7T3P-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
 NIH\_BMAP\_Ret4\_S2 library is a subtracted library,  
 ultimately derived from mouse retina tissue libraries at  
 various stages of development. For a detailed description  
 of the library from which this clone was derived, please  
 visit our web site at [brainest.eng.uiowa.edu](http://brainest.eng.uiowa.edu).  
 TAG\_SEQ=None found"

BASE COUNT 108 a 140 c 119 g 125 t

## ORIGIN

Query Match

8.1%; Score 168.6; DB 11; Length 492;

Best Local Similarity 61.1%; Pred. No. 2.5e-21;

Matches 273; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 1316 agagtgggcagccctcgccatttattctgtgcccgaagtgtatgcacagggaacaggag 1375  
 Db 479 AGGGAGGCCATATCTTTGGCTTCTTATAACCTGCACAGATGCACGCCAGCGGCACAGGGG 420  
 QY 1376 tagtaagatcagcaaaactgctgtgagctttataaaagtctctgtaactaggccaact 1435  
 Db 419 TGATGGCGTCTCTCACACTGCAGTGGAGTTGTTTAAAGATGTGTGAGCGGAGTTCGCT 360  
 QY 1436 gggctgagaaatccctgcagacttacttggctataagatgctgatataagattctctc 1495  
 Db 359 GGTGAGAGACATGATGATGCTCTACACAGCTATTAAAGATGAGACTACATAATGCTGCGAG 300  
 QY 1496 ttgttcagtatgcactgcttcagaaatggggtatgaagttagctcaaaagcaatttcagcat 1555  
 Db 299 TGGTCCAGTACCTCTGCTGGCTGAGCAGGGCTACGAGGTGGCGCAGAGCAACGCGAGCCT 240  
 QY 1556 tcatttggaaataaaaaagggttaacattcttgaataaagaagaagtgtatcccaatggcgc 1615  
 Db 239 TCATCTCTCCAGCAGAGAGAGCAACCACTTGTAGGTGAGAATGAAGCTTATCCCCAGAGCTT 180  
 QY 1616 ttctctatggaatcgagctgccattcaaggcaaatgcttctgtagagtaaaatggag 1675  
 Db 179 TACTGCAATTGGACACAGGGCGCTCCCAAGGTTTACACTGTGGCTAGAAATTAAGCTTTGGAG 120  
 QY 1676 attaccattactatggtatgggactaaagaagactatcaaacagcagccacacactaca 1735  
 Db 119 ACTACCACCTCTATGCTTGGCAGCTGATGTGATTATGAGACCGCATTTATTTCATTACC 60  
 QY 1736 gattgcagccaaataataaccacaacg 1762  
 Db 59 GCCTGGCTTCTGAGCAGCAGCACAGCG 33

## RESULT 12

BF727590

LOCUS

DEFINITION

BF727590 659 bp mRNA EST 08-JAN-2001  
 SWOV3MCAM51C07SK Onchocerca volvulus molting L3 larva cDNA  
 (SL96MLW-Ovml3) Onchocerca volvulus cDNA clone SWOV3MCAM51C07 5',  
 mRNA sequence.

ACCESSION

BF727590

VERSION

EST.

KEYWORDS

ONCHOCERCA VOLVULUS.

SOURCE

ORGANISM

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Onchocerca.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 659)  
 Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.  
 Genes expressed in molting L3 larvae of Onchocerca volvulus  
 Unpublished (1997)  
 Contact: Steven A. Williams  
 Molecular Parasitology  
 Smith College Department of Biological Sciences  
 Department of Biological Sciences, Clark Science Center, Smith  
 College, Northampton, MA, 01063, USA  
 Tel: 4135853826  
 Fax: 4135853786  
 Email: [genomesmith.edu](mailto:genomesmith.edu)  
 Seq primer: pbluescript SK.

## FEATURES

source

Location/Qualifiers

1..659  
 /organism="Onchocerca volvulus"  
 /strain="Kumba, Cameroons"  
 /db\_xref="taxon:6282"  
 /clone="SWOV3MCAM51C07"  
 /clone\_lib="Onchocerca volvulus molting L3 larva cDNA  
 (SL96MLW-Ovml3)"  
 /dev\_stage="molting L3"  
 /lab\_host="XLI-Blue MRF"  
 /note="Vector: Lambda Uni-ZAP XR; Site\_1: Eco RI; Site\_2:  
 Xho I; Filarial nematode parasite of humans. Third-stage  
 larvae, L3, were isolated from infected black flies in







[illegible]

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2002, 20:36:08 ; Search time 53.39 Seconds  
(without alignments)  
8780.848 Million cell updates/sec

Title: US-09-714-882-1  
Perfect score: 2070  
Sequence: 1 atgaagccctgtctctgtt.....ttagaatcaccatggtag 2070

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_NA:\*
- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
  - 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
  - 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
  - 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	146.4	7.1	1124	2	US-08-966-316-11
2	90.2	4.4	571	3	US-08-699-103B-6
3	80.2	3.9	7218	1	US-08-232-463-14
4	64.4	3.1	19124	2	US-08-487-826B-13
5	60.2	2.9	2447	2	US-09-014-969-14
6	59	2.9	240	1	US-08-628-417-6
7	56.2	2.7	6243	2	US-03-056-075-1
8	48	2.3	291	1	US-07-922-723A-7
9	48	2.3	291	1	US-07-799-828C-7
10	48	2.3	291	1	US-08-074-275-7
11	48	2.3	291	1	US-08-480-366-7
12	48	2.3	291	2	US-07-952-277A-7
13	47.8	2.3	4818	3	US-08-817-926-27
14	46.8	2.3	5852	1	US-07-867-106-2
15	46.4	2.2	1117	4	US-09-247-373B-33
16	46.4	2.2	37948	4	US-09-251-645-11
17	46	2.2	636	4	US-08-998-416-1137
18	46	2.2	658	4	US-08-998-416-595
19	45.8	2.2	3763	1	US-07-792-865D-1
20	45.4	2.2	1051	4	US-09-245-041-10
21	44.4	2.1	6671	1	US-08-280-443-1
22	44.4	2.1	6671	1	US-08-457-459-1
23	44.4	2.1	6671	1	US-08-555-678-1
24	44.4	2.1	6671	5	PCT-US95-02275-1
25	44.2	2.1	1813	4	US-09-071-224-3
26	44.2	2.1	6243	2	US-09-056-075-1
27	44	2.1	144	1	US-08-702-344-26

28	44	2.1	19557	5	PCT-US92-06300-1	Sequence 1, Appli
29	43.4	2.1	2223	1	US-08-257-073-4	Sequence 4, Appli
30	43.4	2.1	2946	3	US-08-968-563-6	Sequence 6, Appli
31	43.4	2.1	2946	3	US-08-969-683A-6	Sequence 6, Appli
32	43.2	2.1	1066	1	US-08-157-101A-4	Sequence 4, Appli
33	42.8	2.1	731	1	US-08-451-405A-2	Sequence 2, Appli
34	42.6	2.1	6755	3	US-08-931-999-4	Sequence 4, Appli
35	42.6	2.1	5093	1	US-08-468-036-23	Sequence 23, Appli
36	42.6	2.1	5093	2	US-08-376-843-23	Sequence 24, Appli
37	42.4	2.0	1493	1	US-08-340-820-24	Sequence 24, Appli
38	42.4	2.0	1493	1	US-08-593-535-24	Sequence 24, Appli
39	42.4	2.0	2082	2	US-08-785-310A-2	Sequence 2, Appli
40	42.4	2.0	19124	2	US-08-487-826B-13	Sequence 13, Appli
41	42.2	2.0	376	2	US-08-623-906A-18	Sequence 18, Appli
42	42.2	2.0	454	2	US-08-623-906A-6	Sequence 6, Appli
43	42.2	2.0	837	4	US-08-998-416-288	Sequence 288, App
44	42.2	2.0	1434	2	US-08-903-801-2	Sequence 2, Appli
45	42.2	2.0	1434	4	US-09-295-055-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-966-316-11  
; Sequence 11, Application US/08966316  
; Patent No. 5932445  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Reddy, Roopa  
; APPLICANT: Murry, Lynn E.  
; APPLICANT: Mathur, Preete  
; TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: CA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/08966,316  
; APPLICATION NUMBER: US/08966,316  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0424 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1124 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: LEUKNOT03  
; CLONE: 1880692  
US-08-966-316-11



TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 3.9%; Score 80.2; DB 1; Length 7218;  
Best Local Similarity 5.7%; Pred. No. 2.5e-11;  
Matches 22; Conservative 230; Mismatches 133; Indels 0; Gaps 0;

QY 4 aagccctgtctgttaagagatattgataattcttgggtcaccaataaaactatc 63  
DB 1481 AATTACCTATCTCAAGTACTTAAAGAGATAGAGAATTTGGTACRRRRRRRRR 1422  
QY 64 aagcagagagacataataaaagacaaagaaatgtccacacaggtatcagtg 123  
DB 1421 RR 1362  
QY 124 aacgaatacaacaatatttatcacacatttggaacaaagacatctagtaatc 183  
DB 1361 RR 1302  
QY 184 aataaaagagaaatctctgagaaaaagaaagaaatcaacgtataaataaagga 243  
DB 1301 RR 1242  
QY 244 attcaataaagatatcttgagagaaataagaaatcttcaaaaagcaagcagagaaa 303  
DB 1241 RR 1182  
QY 304 aatttacagtagaggagaccagctatttaagatgggcatacaagttccagcagctc 363  
DB 1181 RR 1122  
QY 364 aaaaacaaacaaagaaagaaag 388  
DB 1121 RR 1097

RESULT 4  
US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827

GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B

FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 3.1%; Score 64.4; DB 2; Length 19124;  
Best Local Similarity 49.4%; Pred. No. 5e-07;  
Matches 167; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 49 acaattaaactatcaagcagagagacataataaaagacaaagaaagaaatgtcacc 108  
DB 15462 AAAATGAAATATAAAAAAATTTTAAATATAAAAAAATAAAAAAAGGAGAAA 15521  
QY 109 acacagggtatcagtgaaacaaatatttatcacacatatttggaacaaagaaaca 168  
DB 15522 AATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATG 15581  
QY 169 tctagtaagttaactaaagagagaaatctcctggagaaagaaagaaatcaacgtaaa 228  
DB 15582 AAAAGATTATCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15641  
QY 229 ataagaataaaggaaattcaaaataaagatatcttgagagagaaataagaatcattaca 288  
DB 15642 AAAAATAAAAAAGAGAGAAAAAATTTAAATATAAAAAAATAAATAAATAAATAA 15701  
QY 289 aagcagcagagaaatattttacagatgaagagagaccgctatttaagatggcctcaag 348  
DB 15702 AAACAAAAAAGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15761  
QY 349 gtctccagcagctctaaagccaaacaaacaaagaaaga 386  
DB 15762 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15799

RESULT 5  
US-09-014-969-14  
Sequence 14, Application US/09014969  
Patent No. 5965397  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallie, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Werberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5951
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-014-969-14

Query Match 2.9%; Score 60.2; DB 2; Length 2447;
Best Local Similarity 53.1%; Pred. No. 2.5e-06;
Matches 128; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 10 ttgtctctgttaagagagatatgtataattcttgggtgcacaaataaaactcaaaagca 69
Db 2207 TTTTGTACTTTAAATGTGACAAATAAACCTTTTGGGAGAAAAAARAAAAA 2266

QY 70 gaggaacataataaagacaaagaagaagaatgtcaccacacaggtatcagtgaaacaa 129
Db 2267 AAAAAA 2326

QY 130 atcaaacataattttcacacatatattggacaaagaacatctagtaattgtaatacaataa 189
Db 2327 AAAAAA 2386

QY 190 agagaaatctctcgtggagaaagaagaatcaacgttaataagaataaaaggaattcaa 249
Db 2387 AAAAAA 2446

QY 250 a 250
Db 2447 A 2447

RESULT 6
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
; ADDRESSEE: DEFENSE COMMAND
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
; CITY: ABERDEEN PROVING GROUND
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21010-5423
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: US/08/628,417
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BIFFONI, ULYSSES J.
; REGISTRATION NUMBER: 39,908
; REFERENCE/DOCKET NUMBER: DAM 398-94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 410-671-1158
; TELEFAX: 410-671-2534
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligodeoxynucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; US-08-628-417-6

Query Match 2.9%; Score 59; DB 1; Length 240;
Best Local Similarity 53.7%; Pred. No. 1.9e-06;
Matches 122; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 52 attaaactatcaagcagagagaaacataataaaagacaaagaagaataatgtcaccaca 111
Db 14 AATAAACTTTAGAAATAATTTTACTAAAAA 73

QY 112 caggtatcagtgacgaatacaacaataattttcacacatatattggacaaagaacatct 171
Db 74 AAAAAA 133

QY 172 agtaatgtaatacaataaaagaagaataatctctcgtggagaaagaagaatacaacgtaaaaa 231
Db 134 AAAAAA 193

QY 232 agaataaaggaattcaaaataaagatatcttgaagagaataaaga 278
Db 194 AAAAAA 240

RESULT 7
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,075
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
```











Db 1090 aaacaaaaaaamaaaaaa 1111

Search completed: April 8, 2002, 21:32:04  
Job time: 3356 sec

